The London School of Economics and Political Science

Managing uncertainty in medicines approval: empirical analyses of non-randomised evidence and regulatory practice

Maximilian Salcher-Konrad

A thesis submitted to the Department of Health Policy of the London School of Economics and Political Science for the degree of Doctor of Philosophy

London, January 2026

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I declare that my thesis consists of 78,260 words.

Statement of co-authored work

This doctoral thesis consists of an introduction, a literature review, three empirical chapters, a discussion, and a conclusion. The introduction, literature review, discussion, conclusion, and one of the empirical chapters are entirely my work. Part of the work presented in two of the empirical chapters was published as co-authored manuscripts in peer-reviewed journals. I confirm that I am fully responsible for the entirety of the work presented in this doctoral thesis other than where I have cited the relevant work of others.

Part of the work presented in chapter 3 was published in *JAMA Network Open.* The published manuscript was co-authored by Dr Mary Nguyen (London School of Economics and Political Science), Dr Jelena Savović (University of Bristol), Prof. Julian P. T. Higgins (University of Bristol), and Dr Huseyin Naci (London School of Economics and Political Science). I conceived the study, led its design and development, collected and analysed data, interpreted the results, and drafted the manuscript. Dr Nguyen supported data collection under my supervision. Dr Naci and Dr Savović provided supervisory guidance on the design of the study, and Prof. Higgins provided supervisory guidance on the statistical analysis. All co-authors reviewed and provided comments on the draft manuscript, and contributed to the final and published version of the manuscript.

Citation: Salcher-Konrad M, Nguyen M, Savovic J, Higgins JPT, Naci H. Treatment Effects in Randomized and Nonrandomized Studies of Pharmacological Interventions: A Meta-Analysis. JAMA Network Open. 2024;7(9): e2436230.

Part of the work presented in chapter 4 was published in *The Milbank Quarterly.* The published manuscript was co-authored by Dr Huseyin Naci (primary Ph.D. supervisor at the London School of Economics and Political Science) and Dr Courtney Davis (King's College London). I conceived the study together with Dr Davis and Dr Naci. I led the design and development of the study, collected and analysed data, interpreted the results, and wrote the first draft of the manuscript. Dr Naci and Dr Davis provided guidance on the study design, data analysis, and interpretation of

findings, reviewed and provided comments on the draft manuscript, and contributed to the final and published version of the manuscript.

Citation: Salcher-Konrad M, Naci H, Davis C. Approval of Cancer Drugs With Uncertain Therapeutic Value: A Comparison of Regulatory Decisions in Europe and the United States. Milbank Quarterly. 2020;98(4): 1219-56.

The work presented in chapter 5 has not been published at the time of writing this thesis. I am planning on submitting the study presented in chapter 5 for publication in a peer-reviewed journal. I conceived the study, led the design and development of the study, collected and analysed data, interpreted the results, and drafted the manuscript.

The papers described above represent the substantive empirical work of this thesis. Over the course of my PhD, I have also been involved as author and co-author in other studies and reviews on issues related to pharmaceutical regulation, study designs and data sources, as listed below in order of time of publication. When drawing on these, I reference them to draw a clear distinction between my thesis and other work.

- *Salcher M. Connecting the dots: putting big data to work for health systems. Eurohealth. 2017;23(1):3-6.*
- *Naci H, Salcher-Konrad M, Kesselheim AS, Wieseler B, Rochaix L, Redberg RF, et al. Generating comparative evidence on new drugs and devices before approval. The Lancet. 2020;395(10228):986-97.*
- *Salcher-Konrad M, Naci H. Unintended Consequences of Coverage Laws Targeting Cancer Drugs. J Law Med Ethics. 2020;48(3):552-4.*
- *Kent S, Salcher-Konrad M, Boccia S, Bouvy JC, de Waure C, Espin J, et al. The use of nonrandomized evidence to estimate treatment effects in health technology assessment. Journal of Comparative Effectiveness Research. 2021;10(14):1035-43.*
- *Cherla A, Mossialos E, Salcher-Konrad M, Kesselheim AS, Naci H. Post-marketing requirements for cancer drugs approved by the European Medicines Agency, 2004-2014. Clinical Pharmacology & Therapeutics. 2022;112(4):846-52.*
- *Davis C, Wagner AK, Salcher-Konrad M, Scowcroft H, Mintzes B, Pokorny AMJ, et al. Communication of anticancer drug benefits and related*

uncertainties to patients and clinicians: document analysis of regulated information on prescription drugs in Europe. BMJ. 2023;380:e073711.

- *Xander NSH, Belleman T, Salcher-Konrad M, Hendrickx A, Chen J, Klein Gebbink AS, et al. Price determinants and pricing policies concerning potentially innovative health technologies: a scoping review. Eur J Health Econ. 2025.*

Acknowledgments

Part of the work presented in chapter 4 has received funding from the European Union Horizon 2020 Research and Innovation Programme (grant agreement No. 779312). Part of the work presented in chapter 5 was supported by funding from Health Action International. The funders had no role in study design, data collection, analysis, interpretation, or publication decisions.

This thesis is the result of my work on pharmaceutical regulation and methodological aspects of comparative effectiveness research over almost a decade. Throughout this period, I was fortunate to receive support in both my private and professional life, as well as critical input and valuable feedback on my research. I cannot name all family members, friends, and colleagues who accompanied me on this journey, but I am deeply grateful to all of you. I would like to thank in particular:

Dr Huseyin Naci, for your careful and empathetic guidance throughout my career, for encouraging me to pursue this PhD, and for setting standards for critical research on pharmaceutical policy which have shaped my research for this thesis and beyond.

Prof. Alistair McGuire, for your continued support throughout my PhD and in my development as a researcher.

Dr Courtney Davis, for your mentorship and the many stimulating discussions on pharmaceutical regulation, without which this thesis would look very different.

Adelina Comas-Herrera, Dr Panos Kanavos, Prof. Martin Knapp, and Prof. Elias Mossialos, for the opportunity to grow as researcher and to learn from you.

Dr Sabine Vogler, for your unwavering support of this PhD, and the entire team at GÖG, for allowing me to focus on finalising this thesis.

Dr Katy Davis, Dr Mary Nguyen, Dr Ishan Rola, and Ker Ro Toh, for supporting me in my sometimes overly ambitious research projects.

Anna-Maria, Arnoupe, Erica, Hikaru, and Olivier, for your deep friendship, many laughs, and for sharing the highs and lows of life in academia.

Marie, for always being there for me, for making me feel special, and for giving me the space and time needed to complete the PhD.

And finally, Romy, for re-calibrating what happiness means, and for giving me the final push to finish this thesis.

Abstract

The introduction of new medicines to the market requires robust evidence about efficacy and safety, traditionally obtained from randomised controlled trials (RCTs). Aiming to accelerate market access, regulatory bodies in Europe (EMA) and the United States (FDA) increasingly accept uncertainty by granting approval based on non-randomised studies and surrogate endpoints. This thesis applies documentary analysis and meta-epidemiological methods to critically examine the methodological assumptions underpinning regulatory acceptance of uncertainty and the regulatory instruments used to manage it in three empirical studies.

First, a meta-epidemiological study of 346 meta-analyses (2,746 studies) found no strong evidence of systematic over- or underestimation of treatment effects in non-randomised studies compared to RCTs overall. However, for a substantial proportion of clinical questions, the two study types led to different conclusions about existence or magnitude of effect, highlighting uncertainties about benefits and risks when substituting RCTs with non-randomised studies. Second, a comparative analysis between EMA and FDA of 21 cancer medicines approvals (2009-2013) with uncertainties in the pivotal trial evidence showed frequent divergence in regulatory outcomes, with one regulator granting full and the other conditional approval requiring confirmatory post-marketing studies or denying approval. When confirmatory studies were imposed, they were frequently delayed and continued to use non-randomised designs and surrogate endpoints. Third, a study of 55 EMA-approved cancer medicine indications (2014–2023) found that single-arm trials increasingly serve as pivotal evidence, yet justifications for accepting methodologically limited evidence were often unclear. For most indications, RCTs appeared feasible, but post-approval RCT evidence only demonstrated clinical benefit on patient-relevant outcomes for a minority of indications.

The findings indicate that reliance on non-randomised studies for regulatory approval introduces uncertainty that may not be resolved post-approval. Regulators should leverage their position as gatekeepers to incentivise robust evidence generation as part of research and development programmes.

Table of Contents

Table of Figures

Table of Tables

Table of Boxes

Abbreviations

AA	Accelerated Approval
ALL	Acute lymphoblastic leukaemia
BTD	Breakthrough therapy designation
CI	Confidence interval
CrI	Credible interval
CHMP	Committee for Medicinal Products for Human Use
CLL	Chronic lymphocytic leukaemia
CMA	Conditional Marketing Authorisation
CML	Chronic myeloid leukaemia
DARWIN EU	Data Analysis and Real World Interrogation Network
EFS	Event-free survival
EGFR	Epidermal growth factor receptor
EHR	Electronic health record
EMA	European Medicines Agency
EPAR	European Public Assessment Report
ESMO	European Society of Medical Oncology
ESMO-MCBS	ESMO-Magnitude of Clinical Benefit Scale
EU	European Union
FDA	Food and Drug Administration
FDCA	Federal Food, Drug, and Cosmetic Act
FTD	Fast track designation
GMP	Good Manufacturing Practice
HL	Hodgkin lymphoma
HR	Hazard ratio
HTA	Health technology assessment
ICH	International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use
IQR	Interquartile range
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
IR	Incidence ratio
JCA	Joint clinical assessment
JSC	Joint scientific consultation
MA	Meta-analysis
NICE	National Institute for Health and Care Excellence
NRS	Non-randomised study
OR	Odds ratio
ORR	Objective (or overall) response rate
OS	Overall survival
PFS	Progression-free survival
PICO	Participants, intervention, comparator, outcome(s)
PMO	Post-marketing obligation
PRAC	Pharmacovigilance Risk Assessment Committee
PRIME	Priority Medicines
PTCL	Peripheral T-cell lymphoma
QALY	Quality-adjusted life year

R&D	Research and development
RCT	Randomised controlled trial
ROR	Ratio of odds ratios
RR	Risk ratio
RWD	Real-world data
RWE	Real-world evidence
sALCL	Systemic anaplastic large cell lymphoma
SAT	Single-arm trial
SE	Standard error
SMD	Standardised mean difference
SmPC	Summary of product characteristics
TKI	Tyrokinase inhibitor
US	United States (of America)

1. Introduction

1.1. Regulating market entry of new medicines

The introduction of new medicines to the market is heavily regulated by legislation. In the European Union (EU) and the United States (US), pharmaceutical manufacturers have to obtain approval for new products from regulatory bodies (European Medicines Agency, EMA, and the Food and Drug Administration, FDA, respectively), before they are allowed to enter the market. The mission of these agencies is to protect public health and to ensure the availability of high quality, safe and effective medicines for citizens.[1, 2] In the US, current legislation dates back to the 1962 Kefauver-Harris amendments, which were introduced partly in response to the high-profile case of thalidomide, which caused birth defects in some newborns. This marked a turning point in medicines approval legislation in the US: in contrast to the previous focus on safety and post-marketing regulatory actions, new medicines now had to be approved prior to market entry, and they had to demonstrate both safety and efficacy through "well controlled clinical trials".[3, 4] The Kefauver-Harris amendments changed medicines regulation in the US, but through the leading role of the US regulatory agency, FDA, also impacted on regulatory practices in other countries and regions. In the EU, the mandate for regulation of new medicines evolved from a national competence of the EU member states towards a shared and common standard across the EU, operationalised through the EMA. The EMA, which is responsible for new medicines authorised through the centralised procedure, was established in 1995, when an international consensus already existed that new medicines need to be proven both safe and effective before they become available to patients. Medicines regulatory bodies exist in other jurisdictions as well, but this thesis focuses on EMA and FDA since they occupy leading roles in shaping evidence standards and act as gatekeepers to some of the most important markets for pharmaceutical companies.

This thesis contributes to the multidisciplinary field of regulatory science, i.e. research that aims to inform regulatory decision-making.[5] Regulatory science in general has been defined as the body of knowledge and tools used to assess products within a regulated industry,[6] and the evaluation of the impact of regulation.[7]

Within this broad field, regulatory science as applied in the pharmaceutical sector is particularly concerned with the development of *new* tools, standards, and approaches for evaluating medicines within a regulatory framework,[8] the scientific and non-scientific factors shaping the framework, and its outcomes.[5] The scope of regulatory science includes the methods and data used to assess evidence on the efficacy, safety, and quality of medicines, as well as setting standards along the stages of the product life cycle, from developing new medicines to post-market surveillance, and the systematic evaluation of regulatory tools, pathways, and decision-making processes. Regulatory science therefore sits at the interaction of science and policy, representing the tools and approaches used to promote public health and safety.[6] Applying the multidisciplinary perspective of regulatory science, this thesis contributes to both the methods used to assess new medicines and the policy question of regulatory acceptance of uncertainty in clinical evidence.

The scope of regulatory assessment is to determine whether a new medicine provides more benefit than harm to patients. Market approval represents the first regulatory step in determining which therapies become available in clinical practice. While evidence standards exist, regulators commonly have to manage uncertainty in the clinical evidence about a medicine's benefits and harms. When left unaddressed, such uncertainty impacts downstream actions and decision-making, including clinical guidelines, reimbursement decisions, and prescribing behaviour. Evidence standards applied in regulatory assessments – focusing on the balance between benefits and harms of medicines – therefore have implications not only for regulators, but also for patients, clinicians, payers, and other decision makers. By contrast, health technology assessment (HTA) bodies, payers, and medical societies are interested in the added therapeutic benefit compared to existing alternatives. Despite these differences in the assessment scope, the evidence generated for regulatory approval is also used to inform reimbursement and treatment decisions, as this is typically the most methodologically robust evidence available.

In addition to efficacy and safety, regulatory approval includes an assessment of the pharmaceutical quality of the product. This ensures that authorised medicines are manufactured, stored, and delivered to a high standard and includes, among other

aspects, adherence to Good Manufacturing Practice (GMP). Quality aspects are not considered in this thesis.

1.2. Evidence standards for medicines regulatory approval

1.2.1. RCT evidence as gold standard for medicines regulatory approval

Market approval applications have historically relied on data from clinical studies that allow an internally valid assessment of meaningful benefits for patients. Randomised, controlled trials (RCTs) represent the gold standard in evaluation research and, if properly conducted, allow unbiased estimation of the causal effect of an intervention.[9] Using a random sequence, study participants are allocated to either an intervention or a control group, thereby theoretically achieving a distribution of observed and unobserved participant characteristics that is not systematically different in one group vs. the other. This design, along with a protocol-driven controlled environment, ensures internal validity: the observed treatment effect can be attributed to the experimental treatment.[10] Their internal validity places RCTs (and their synthesis through meta-analyses and systematic reviews) at the top of traditional evidence hierarchies.[11, 12] Randomised allocation of participants to intervention and control groups by itself, however, does not suffice to obtain unbiased estimates of treatment effects. Meta-epidemiological research has empirically identified potential sources of bias in RCTs. Risk of bias due to an improper randomisation process, deviations from intended interventions, missing outcome data, issues with outcome measurement, and selected reporting of results accordingly all feature in the widely used Cochrane Collaboration's risk of bias tool for assessing RCTs.[13] RCTs forming the basis for regulatory approval of new medicines are not immune from these biases. A study of all pivotal RCTs for new cancer medicines approved by the EMA from 2014 to 2016 found that 49% were at high risk of bias.[14]

Non-randomised studies, on the other hand, are at increased risk of bias due to the lack of random allocation of study participants to treatment and control groups. Different biases can feature in non-randomised studies. The ROBINS-I tool for assessing risk of bias in non-randomised studies, developed by members of the Cochrane Collaboration, lists three types of bias in non-randomised studies which are distinct from RCTs: bias due to confounding through participant characteristics that are

predictive of treatment assignment; bias in the selection of participants into the study based on characteristics that are associated with both intervention status and outcome; and bias in the classification of interventions.[15] The increased risk of bias in non-randomised studies is acknowledged by regulators.[16] For example, the EMA states in guidance on clinical trials that "[i]nterventional studies, and in particular randomised studies, play a central role in drug development, as they can better control biases."[17, p15] The FDA also considers RCTs to generally provide the most robust evidence on clinical benefits of new medicines.[18, 19] Nonetheless, regulators are afforded some flexibility in applying evidence standards and may therefore also consider other study designs as the basis for approval decisions, including non-randomised studies. Importantly, non-randomised studies can provide evidence that RCTs cannot. Regulators therefore commonly rely on non-randomised studies to inform regulatory actions across the product life cycle, including in the understanding of the natural history of disease, relevant patient populations and their health care utilisation patterns, for determining appropriate designs for RCTs and other clinical trials, and for monitoring the safety and use of a newly authorised product after approval.

Challenges in conducting RCTs

While their methodological advantages for obtaining internally valid estimates of treatment effects are widely accepted, RCTs have long been criticised for being costly, more time-consuming, and more difficult to conduct than non-randomised studies. For example, the recruitment of a sufficiently large patient population for a clinical trial can pose a serious problem for the conduct of a study,[20, 21] potentially delaying results on the efficacy of a new treatment. Recruitment for a RCT may be particularly challenging when there is no established treatment and participants are reluctant to be randomised to receiving placebo rather than a promising – although at this point not proven effective – new treatment. From the perspective of the trial sponsor and treating physicians, there is also an ethical challenge in areas of unmet medical need when a new treatment promises significant improvements for patients' lives based on early data. Regulators recognise these challenges and allow flexibility in relation to the study designs acceptable for regulatory approval. In its guidance on demonstrating clinical benefits, the FDA states that "[…] a sponsor should consider the specific clinical circumstance, including the severity of the disease, unmet medical need (e.g., whether

there is available therapy), the rarity of the disease, and whether it is feasible and ethical to conduct a randomized concurrently controlled superiority trial."[18, p14]

Another criticism of RCTs relates to their external validity: the restrictive inclusion/exclusion criteria of RCTs might lead to problems in extrapolating findings to other patient groups that have not been included in the trial, delaying access for patients for whom the treatment might work.[22] However, while external validity is sometimes presented as an important advantage of non-randomised studies compared to RCTs, both study designs can create issues for the generalisability of findings. In a reflection paper, the EMA highlighted issues with external validity outside of RCTs. While trial populations may be systematically different from the target patient population in both randomised and non-randomised studies, heterogeneous treatments effects (such as a better response among patients with a specific biomarker) are more likely to result in biased estimates of the true treatment effect in non-randomised studies where the observed effect cannot be reliably contextualised against a concurrent comparator.[23] There are also efforts to make RCTs better reflective of the real world and to integrate them into clinical practice to enhance their feasibility as so-called pragmatic trials or pragmatic RCTs.[24] These study designs aim to combine the methodological rigour of RCTs with better reflection of clinical practice and have been conducted in different therapeutic areas.[25]

1.2.2. Increasing interest in the use of non-randomised studies

Interest in the use of non-randomised studies for medicines regulatory approval has intensified over the past two decades. While non-randomised studies have always featured in regulatory decision making across the product life cycle, this has traditionally been limited to a supporting role in relation to approval decisions. For example, important roles for observational studies and databases include the characterisation of disease epidemiology and natural history of disease, quantification of unmet medical need, and assessing the feasibility and informing the design of clinical trials. Once a product is authorised, observational post-authorisation safety studies take a prominent role in pharmacovigilance (e.g. registries for documenting adverse events, monitoring off-label use and adherence).

However, the role of non-randomised studies has since evolved, and there is increased interest in their role for determining clinical efficacy both in the pre- and post-marketing setting.[26, 27] This includes non-randomised studies submitted as pivotal evidence for regulatory approval, the use of observational data to contextualise results from single-arm trials, and observational post-marketing studies.[26-33] Senior FDA officials have discussed the role of non-randomised studies using real-world data for regulatory decisions.[34] This included using such data as external controls in situations where a RCT is not feasible (due to the rarity of disease, or no alternative treatment being available to be randomised to) or ethical (due to early evidence indicating a highly effective new treatment), but also for conducting prospective or retrospective observational studies.

Regulators maintain that RCTs will continue to be the gold standard for medicines regulation and that non-randomised studies can provide important complementary evidence, including in situations when RCTs are not feasible or ethical.[34, 35] However, empirical evidence shows an increase in the use of non-randomised studies to provide evidence on efficacy as the basis for approval decisions.[29, 36, 37] Moreover, research on regulatory science and pharmacoepidemiology shows a clear interest in not only complementing – but substituting – RCTs with observational studies.[38] The interest therefore appears to lie not only in utilising the advantages of non-randomised studies as complementary evidence to RCTs, but to use them as primary source of evidence on clinical benefits of new medicines.

Drivers for interest in real-world data

Regulatory bodies, including the EMA and FDA, have actively promoted the use of non-randomised studies. Both agencies have published frameworks on the use of observational (or "real-world") data for regulatory decision making.[35, 39] At the EMA, this included the implementation of a federated data network for real-world data which can be used to conduct studies across different regulatory use cases, including post-marketing studies on safety and medicines utilisation, but also studies on disease epidemiology and effectiveness.[40, 41]

A potential driver for increased interest in and reliance on non-randomised studies are methodological advances in the analysis of observational data and available adjustment methods to allow causal statements about the efficacy of a treatment to be

made. In 1965, Sir Austin Bradford Hill famously suggested a set of nine criteria that, if fulfilled, make causal statements drawn on the basis of observational data more likely to be true.[42] Since then, the discussion has moved on to causal inference and the notion of the counterfactual, where methods have been developed to account for confounding factors in observational studies.[43] Analytical methods with the aim of obtaining unbiased treatment estimates from non-randomised studies include propensity score methods and instrumental variables, among others.[44-46] Propensity score methods in particular have become relatively common tools to adjust treatment effects when evaluating treatments in non-randomised studies.[47] In addition to advanced statistical methods, the use of the "target trial" framework for emulating RCTs using observational has attracted interest for regulatory decision making.[48, 49]

In parallel to methodological advancements, health systems are also increasingly interested in leveraging the vast data they generate and collect. This idea is often conceptualised as "learning health systems", i.e. health systems that continuously generate data and analyse them to improve the care they provide.[50] Learning health systems use the data from clinical practice to generate evidence and thus integrate clinical and research infrastructures. This concept has direct implications for the regulation of the medicines market, as it provides opportunities for evidence generation outside the traditional clinical trial model.[51]

A range of data sources exist which provide information on patients, their characteristics, and health outcomes, but were not specifically designed to provide research data. This includes administrative data such as electronic health records (EHR) and claims data, as well as clinical registries (although registries may also be set up for research purposes). Regulators are interested in making more systematic use of these data. For example, senior FDA officials have cited the availability of EHRs as a key driver for recent interest in using these data to inform regulatory decisions.[34] Registries (a registry is defined as "organised system that collects uniform data (clinical and other) to identify specified outcomes for a population defined by a particular disease, condition or exposure")[52, p4] also have attracted interest from regulators in being used more systematically for medicines evaluation.[53]

Collectively, data that are collected outside traditional clinical trial settings have been labelled as "real-world data" (RWD). These can include any routinely collected

data relating to health outcomes or the delivery of interventions.[41] "Real-world evidence" (RWE) is the evidence derived from real-world data and is, by definition, observational. However, not every non-randomised study generates real-world evidence: interventional non-randomised studies (e.g. single-arm trials or multi-arm trials without randomised allocation to treatment groups), are still clinical trials and are therefore excluded from the definition of real-world evidence even though they are non-randomised. Nonetheless, part of the interest in real-world evidence for regulatory decisions is directly related to such non-randomised clinical studies for two reasons:

- Firstly, real-world evidence can be used to compare treatment effects observed in non-randomised studies through external control groups. Either done explicitly (through formal statistical comparison, potentially adjusting the external control to better match the patients enrolled in the interventional study) or implicitly (through a "ballpark" comparison of treatment effects observed for other treatments), real-world evidence can provide some contextualisation of data from non-randomised studies without a control group (i.e. single-arm trials). An example of this use of real-world data is the approval of avelumab for treatment of metastatic Merkel cell carcinoma. Approval was granted by the FDA in 2017 based on an uncontrolled phase II trial; data from that trial were compared to a historical control group sourced from EHR data.[54]
- Secondly, real-world evidence holds a promise of addressing remaining uncertainties about the clinical benefits of new treatments in the post-marketing setting. Such uncertainties are particularly prevalent for medicines approved based on early data (such as non-randomised phase II trials). On the condition of generating valid effect estimates, real-world evidence could therefore provide confirmatory evidence on clinical benefits after early approval.

A final driver for growing interest in the use of non-randomised studies may therefore lie in the desire to accelerate market access for new and promising medicines. Since preliminary evidence often takes the form of non-randomised studies (e.g. single-arm phase II trials), regulators are interested in understanding the circumstances under which they can interpret data from non-randomised studies to establish a positive

benefit-risk ratio. When deemed reliable, non-randomised studies can therefore lead to earlier market approval. This is particularly relevant for approvals in areas of unmet need where RCT evidence may not be expected due to feasibility or ethical issues. The relationship between non-randomised evidence and timelines for regulatory approval are described in more detail in section 1.5.

1.3. Uncertainty about clinical benefits and trade-offs between evidence standards and early access

The research and development process for new medicines takes many years and includes pre-clinical studies, testing in healthy volunteers and clinical studies in patients.[55] Generating a robust evidence package that addresses all relevant questions about a new product's benefits and harms takes time, as new treatments pass through different stages of research and development. Traditionally, robust evidence packages included large RCTs in phase III, which take on average between two and four years to complete.[56] Regulators are therefore facing trade-offs between the certainty of the evidence and making new and efficacious treatments accessible to patients as quickly as possible. Through their approval decisions, regulators provide a measure of acceptable uncertainty at the time of marketing authorisation.

Regulatory bodies have the mandate to protect public health. This includes safeguarding patients from harmful medicines, but also to ensure that medicines are effective. Legislation in the EU and the US ties marketing authorisation to a favourable benefit-risk profile and requires this to be demonstrated through adequate evidence. This requires a delicate balance between the certainty of evidence about the benefits and harms of new medicines and the health impact on patients. While regulators recognise the methodological advantages of RCTs (see section 1.2 for a discussion of evidence standards for regulatory approval), guidance documents by both EMA and FDA also show some flexibility about the type of evidence that can be considered adequate. For example, the FDA states the following:

> *"FDA must reach the conclusion that there is substantial evidence of effectiveness to approve a drug; however, the degree of certainty supporting such a conclusion may differ, depending on clinical circumstances (e.g., severity and rarity of the disease and unmet medical*

need). This reflects the longstanding awareness that, in certain settings, a somewhat greater risk (compared to placebo-controlled or other randomized superiority trials) of false positive conclusions – and therefore less certainty about effectiveness – may be acceptable, when balanced against the risk of rejecting or delaying the marketing of an effective therapy […]."[18, p14]

While the use of less robust evidence results in uncertainty about the magnitude or even existence of a new medicine's clinical benefit at the time of market entry, this is accepted by regulators in order to facilitate patient access to a new treatment. An increase in regulatory acceptance of uncertainty may reflect a change in how regulators interpret their role. In a 2018 strategic reflection paper, the EMA positioned itself as supporting pharmaceutical innovation, placing innovation and patient access to medicines at the heart of its vision: "To underpin its mission of protecting human health, EMA must catalyse and enable regulatory science and innovation to be translated into patient access to medicines in evolving healthcare systems."[57, p2] Similarly, the FDA in its strategic plan for regulatory science emphasises innovation and facilitating access to medicines: "FDA will advance regulatory science to speed innovation, improve regulatory decision-making, and get products to people in need."[58, p6] These statements show that regulators aim to position themselves not as gatekeepers but as facilitators of innovation and partners in making new treatments available to patients in need.[59, 60]

The trade-offs between uncertainty about clinical benefits and early access to new treatments require explicit consideration. Expedited approval pathways, as described in section 1.4 below, aim to balance the risks and benefits of early access based on less certain data; however, they have attracted critical attention from scholars due to concerns about the study designs and endpoints used in both pivotal trials and confirmatory studies, as well as the delayed or incomplete confirmation of clinical benefits post-approval.[55, 61, 62] The resulting uncertainty about the benefits and harms of new medicines affect patients, prescribers, and health systems as they face challenges in deciding whether new treatments should be used on the basis of limited information.

1.4. Expedited approval pathways

The challenging task of regulators to balance the protection of public health by ensuring new medicines are safe with the requirement to make effective treatments accessible as quickly as possible has been labelled the "evidence versus access conundrum".[63, 64] In response, new mechanisms in the regulatory decision-making process were introduced which seek to shorten the entire process of medicines development and regulatory review.[65] In Europe, these include accelerated assessment, conditional marketing authorisation, approval under exceptional circumstances, orphan designation, and PRIME. In the US, similar pathways exist (usually pre-dating the European routes): priority review, accelerated approval, fast track designation, orphan designation, and breakthrough therapy.[3, 55, 66, 67] Table 1 provides an overview of these pathways in the EU and the US.

These schemes, commonly referred to as "expedited approval pathways", aim to accelerate either the regulatory review process or the research and development process, or both. Both play a role in determining time to market access for new medicines. The average duration of clinical research and development (i.e. in phases I, II, and III) for new medicines is approximately six to eight years,[56] and the median time from submission of the application to the final regulatory decision is around one year.[68, 69]

Efforts to accelerate the time spent on generating evidence can include a change to the evidence standards usually required for regulatory approval. For example, accelerated approval in the US can be granted when efficacy is demonstrated for a surrogate endpoint which is considered likely to predict benefit (in contrast to the use of more established surrogate endpoints for regular approval), and in Europe, conditional marketing authorisation can be granted based on early evidence that needs to be confirmed through post-marketing studies.[67] Figure 1 provides a schematic illustration of regulatory approval through regular or conditional approval pathways in relation to the evidence generated: in contrast to regular approval, conditional approval may be granted on the basis of phase II trial data, which may then be required to be confirmed through a phase III trial or other studies taking place in the post-marketing setting.

While expedited approval pathways have indeed led to shorter regulatory review times,[68] critical voices have raised concerns over potential lowering of the evidence standards expected from the assessment of new medicines before market entry.[70-73] Expedited approval pathways may impact on evidence generation patterns more broadly, leading to a landscape of small and low quality studies.[62] There are also questions whether expedited pathways do indeed accelerate patient access. In Europe, medicines approved through the conditional marketing authorisation pathway did not have shorter time to approval than products with regular approval, although in the US, accelerated approval did reduce time to market.[74]

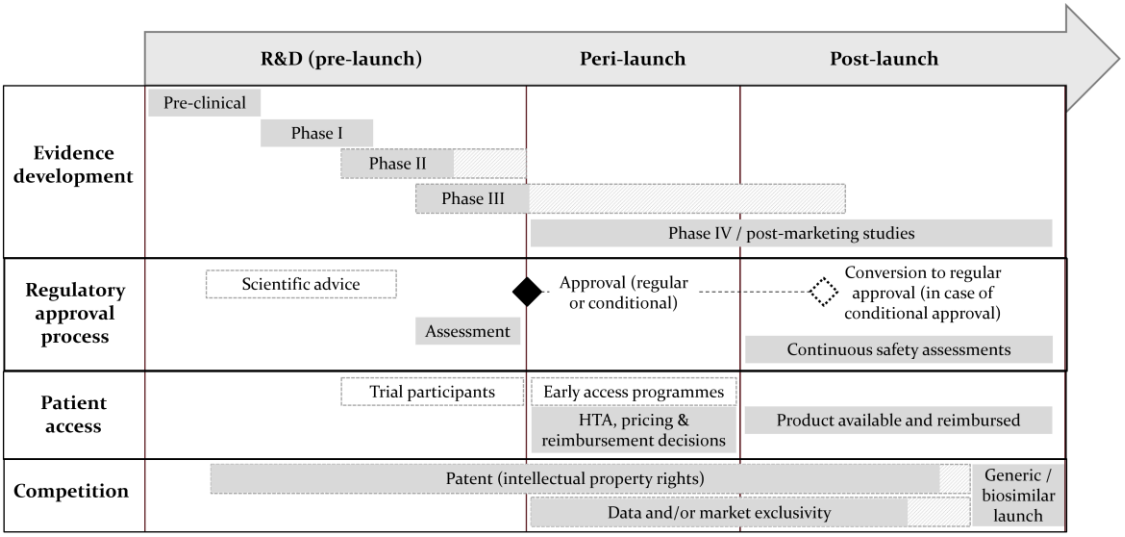
Table 1: Expedited approval pathways in the EU and US

EU (EMA)	US (FDA)
Accelerated assessment: reduces the assessment time from 210 to 150 days for products that are expected to be of major public health interest in particular from the viewpoint of therapeutic innovation.	**Priority review:** a designation for products that treat serious or life-threatening conditions and that promise significant improvement in safety or effectiveness compared to existing treatments. Evidence for such superiority would typically be expected to come from RCTs, but the FDA has also stated that observational study designs, such as historical controls, can be sufficient.[75] Priority review does not formally change evidence standards, but shortens the time available for assessing new medicines.[76]
Conditional marketing authorisation (CMA): regulatory pathway for approval based on limited clinical data.[67] CMA allows early approval for products intended for the treatment, prevention or medical diagnosis of seriously debilitating or life-threatening diseases based on data that would otherwise not be considered sufficient for approval, conditional on confirmatory evidence being produced in the post-marketing setting. This requires four criteria to be met, including a positive benefit-risk	**Accelerated approval (AA):** regulatory pathway for approval based on limited clinical data.[67] AA allows early approval for medicines treating serious or life-threatening diseases based on data that would otherwise not be considered sufficient for approval, conditional on confirmatory evidence being produced in the post-marketing setting. Specifically, AA allows submission on the basis of surrogate endpoint that are "reasonably likely" to predict clinical benefit.

ratio, an unmet medical need being addressed, the ability of the developer to provide confirmatory evidence, and the benefit of immediate availability of the product outweighing the risks due to the incomplete data package. CMAs are valid for one year and subject to annual review.	
Approval under exceptional circumstances: regulatory pathway for approval based on limited clinical data.[67] Approval under exceptional circumstances allows submission of incomplete data when it is considered unethical or impossible to collect additional data, e.g. due to the disease being rare.	No equivalent pathway in the US.
Priority Medicines (PRIME): a scheme that aims to engage manufacturers and regulators in an early dialogue about clinical trial design to ensure that submitted evidence is appropriate for marketing authorisation for medicines that address unmet clinical need or that offer a major advantage over existing alternatives. PRIME designation allows the use of accelerated assessment and CMA.[67]	**Fast Track designation (FTD):** applications can be submitted under the FTD if the drug has the potential to addresses unmet clinical need for serious or life-threatening conditions. The designation allows more frequent interactions between the drug manufacturer and the FDA, including discussions on study design. Evidence requirements can vary based on the development stage of the drug and can include clinical and non-clinical data.[75] FTD includes the possibility of receiving marketing authorisation on the basis of a single phase II trial.[55]

	Breakthrough Therapy designation (BTD): a designation that can be awarded if preliminary evidence (typically phase I or phase II trials) indicates a substantial improvement compared to existing treatment (or placebo, if no treatment exists). BTD allows the manufacturer to meet with the FDA more regularly. The designation does not change evidence standards, as products with BTD are approved through standard FDA approval routes,[76] but products with the designation have substantially shorter development and review times.[77]
Orphan designation: offers the manufacturer market exclusivity for 10 years as an incentive to invest in areas with small patient numbers. Medicines meeting the criteria for orphan designation (intended for treatment of life-threatening or seriously debilitating conditions that affect no more than 5/10,000 people) are eligible for CMA (see above).	**Orphan drug designation:** grants the marketing authorisation holder market exclusivity for seven years. This designation does not formally change evidence standards.[55]

Figure 1: Selected events and processes in the pharmaceutical product life cycle



Legend: Figure shows selected events and processes in the pharmaceutical product life cycle across three phases (pre-launch, peri-launch, and post-launch) and for four dimensions (evidence development, regulatory approval process, patient access, and competition). Timelines are for illustrative purposes only. Dotted lines and boxes indicate scope for changes in timelines and events depending on regulatory approval processes. Not displayed are iterative evidence generation cycles across different indications for the same product.

Abbreviations: HTA, health technology assessment; R&D, research and development.

1.5. Evidence generation and the product life cycle

1.5.1. Pharmaceutical product life cycle

When analysing medicines regulatory practices and their impacts, it is instructive to consider the pharmaceutical product life cycle (Figure 1). Broadly, this distinguishes three phases.

(1) In the pre-launch phase, research and development of a new product take place. At this stage, regulatory bodies may provide scientific advice to medicines developers. Towards the end of this phase, the evidence generated is submitted to the regulatory body and assessed to determine whether marketing authorisation can be granted. Prior to marketing authorisation, typically only patients enrolled in clinical trials have access to the treatment. Intellectual property rights ensure that developers can bring their product to market without competition.

(2) Upon receiving marketing authorisation, the peri-launch phase starts where third-party payers (national health systems and insurers) determine whether the new product should be reimbursed. This phase is particularly relevant for Europe and other jurisdictions with price regulation for pharmaceuticals and national health systems or social health insurance that require an assessment of the comparative benefits of the new product in relation to existing treatments. HTA determines whether the new treatment provides value-for-money. During this time, patients may receive treatment through early-access programmes, but regular access for all patients is typically contingent on a positive reimbursement or coverage decision. In the US health system with a greater reliance on private health insurance, pricing is less regulated compared to Europe and HTA plays a less important role for informing reimbursement decisions; for cancer medicines, legal provisions exist that mandate reimbursement, leaving little room for insurers to base coverage decisions on value assessments.[78-80]

(3) In the post-launch phase, the product is available for patients. From the regulatory perspective, the focus shifts towards continuous monitoring of safety (pharmacovigilance). Additional post-marketing studies may be required to address open questions. For conditional approvals, evidence generation to confirm clinical benefit continues into the post-launch phase. Depending on the timing of patent filing and regulatory approval, the period under which the product is protected from competition ends and generic or biosimilar versions can enter the market and broaden patient access.

The description above and displayed in Figure 1 is intended to provide a high-level overview of different phases in the product life cycle that are of relevance to this thesis and does not aim to provide an exhaustive description.

1.5.2. Evidence development throughout the life cycle

As new medicines pass through the different stages of research and development, uncertainty about their benefits and harms is expected to decrease with each new phase. After pre-clinical research and development, investigational products typically undergo clinical testing in three phases. Phase I trials are conducted in healthy

volunteers and aim to assess safety and dosage; these trials are a necessary pre-requisite for testing the new treatment in patients. Phase II trials assess safety and efficacy in patients, and phase III trials provide confirmatory evidence on a new treatment's benefits and harms. Phase III trials are typically conducted as randomised trials comparing the new treatment to either placebo or the current standard of care, although they can also be conducted as single-arm trials.

Uncertainty in non-randomised phase II trials

Phase II trials can be also conducted as comparative trials (including comparisons of different doses of the new treatment or comparisons against placebo) or single-arm trials. In the latter, only endpoints can be measured with confidence that would not be expected to be observed without activity of the treatment under investigation.[23] A typical example in oncology would be tumour response: a tumour is highly unlikely to naturally shrink over time; if the natural progression of the cancer is known and tumour shrinkage is observed in a non-comparative trial of a new treatment, the response can therefore be ascribed to the treatment. However, tumour response is a surrogate endpoint. When not validated as a robust surrogate for clinically meaningful outcomes, evidence on tumour response carries uncertainty about the clinical benefit of the treatment.[81] To confirm clinical benefit, the effect of the new treatment on patient-relevant endpoints such as overall survival or quality of life should be demonstrated. This requires a control group to provide the counterfactual. In a RCT, any differences between the intervention and control groups are due to chance alone, minimising bias arising from a systematically different composition of the two groups. As described in section 1.2.1, non-randomised studies, on the other hand, are at risk of bias due to confounding (i.e. participant characteristics are predictive of treatment assignment and are also associated with the outcome), bias in the selection of study participants (i.e. participant characteristics are predictive of participation in the study and also associated with the treatment received and the outcome), and bias in the classification of interventions.[15] A higher risk of bias compared to properly conducted RCTs exists for concurrently controlled non-randomised studies, but also for studies with external (or historical) control groups. In case of the latter, internal validity of the comparison of outcomes is also jeopardised through a potentially different environment for the external control group, e.g. due to different treatment standards at the time of

data collection (in case of historical cohorts) or in a different region or country. Whether conducted as single-arm trials without any comparator or as non-randomised study with external controls, these study designs are at risk of bias and therefore introduce uncertainty about the true treatment effect.

Timelines for availability of confirmatory evidence

As shown in Figure 1, the timelines of when data from phase II and phase III trials are available can change depending on the type of regulatory approval. Approval is typically based on data from phase III trials. However, some expedited approval pathways also allow approval based on earlier data (Table 1): in conditional approval pathways (conditional marketing authorisation in the EU, accelerated approval in the US), initial marketing authorisation is granted based on preliminary data that would not usually be considered sufficient for regular approval, such as data from a phase II trial. This would then need to be confirmed through additional studies in the post-marketing setting. In the traditional model for conditional approval pathways, confirmatory studies are phase III trials which would be underway at the time of (early) marketing authorisation.[82] An example for early (conditional) approval based on data from a non-randomised phase II study is trastuzumab deruxtecan for treatment of breast cancer after prior failed therapies, which received conditional marketing authorisation by the EMA in 2021.[83] Clinical benefit was later confirmed through a RCT demonstrating superiority on both progression-free survival and overall survival against standard treatment.[84] In many other cases, confirmatory evidence from RCTs measuring patient-relevant outcomes is lacking.[61, 85]

1.5.3. Product availability and protection from competition

The timelines for evidence generation can also be discussed against the point at which patients can access new medicines, and the period during which medicines developers are protected from competition.

As can be seen from the schematic illustration in Figure 1, regulatory approval based on preliminary data can enable earlier patient access. After regulatory approval, HTA may be conducted to determine routine availability of a new medicine in a given market (i.e. whether the product is reimbursed), and the price is set in jurisdictions with relevant applicable pricing and reimbursement processes. For new medicines

addressing an unmet medical need, the latter often occurs through negotiations between the company and the payer since no reference treatments exists. While approval based on early data accelerates the time at which a new treatment is available, this comes at the cost of uncertainty. The uncertainty is relevant for patient access because it creates challenges for HTA bodies, payers, and physicians in assessing its added therapeutic value, and therefore creates uncertainty about whether and for whom the new treatment should be used.

The illustration also shows that early market access expands the effective period of market exclusivity for the company. The effective period of market exclusivity is determined by both intellectual property rights (i.e. patents) and data or market exclusivity rights associated with marketing authorisation.[86] Data exclusivity is the period during which data relating to marketing authorisation of a new product cannot be used by manufacturers of competitor products to prepare marketing authorisation applications for generic or biosimilar versions of that product, and market exclusivity is a period where such data can already be used by competitors but their products are not yet allowed to enter the market. These periods are linked to the date of marketing authorisation.[86] Intellectual property rights, on the other hand, are independent of whether a product received regulatory approval. Yet, the time at which marketing authorisation is granted still matters for the period during which the company has monopoly protection. In the EU and the US, the patent duration is 20 years from the date of filing. However, patent filing typically occurs early in the research and development process. The effective period before generic competition occurs is between 12 to 15 years.[87] Shifting the time of marketing authorisation forward in relation to the patent filing date therefore extends the period during which intellectual property rights apply. This also results in a period during which the product is authorised, but where uncertainty persists over its clinical benefit.

1.6. Key assumptions about the role of non-randomised studies for accelerating market access

There is growing interest in using non-randomised studies for regulatory approval to accelerate market access for new medicines, as described in section 1.2.2. Expedited regulatory pathways can act as tools for accepting evidence from non-randomised

studies, and may also play a role in further promoting their use, including as pivotal evidence at the time of approval and as confirmatory evidence in the post-marketing setting, as described in sections 1.4 and 1.5. Approval based on non-randomised studies represents a deviation from the long-held evidence standard for medicines marketing authorisation. However, flexibility in regulatory standards may be appropriate in some cases to ensure patients receive the treatments they need.

Such flexibility is motivated by the expectation that acceptance of non-randomised studies for regulatory approval leads to accelerated market access for medicines with a net positive impact on public health. Indeed, in the EU, the conditional marketing authorisation pathways requires that early availability of a new treatment offsets the potential risks posed through the preliminary nature of the evidence. Whether a positive benefit-risk balance exists has to be assessed with the evidence available. Unless robust evidence on the clinical benefits of a treatment becomes available, establishing such positive impact on public health through expedited approval is therefore fraught with uncertainty.

The case of brentuximab vedotin, an antibody-drug conjugate for lymphoma treatment, illustrates the relationship between uncertainty and study design (see section 4.3.5 for a brief case study on the approval of brentuximab vedotin in the EU and the US): in 2012, the EMA granted conditional approval for brentuximab vedotin for treatment of Hodgkin lymphoma and systemic anaplastic large cell lymphoma on the basis of interventional non-randomised studies (clinical trials without a control group).[88] While the evidence from non-randomised pivotal studies was considered sufficient to establish a positive benefit-risk ratio, the methodological limitations (i.e. lack of control groups) meant that further studies were deemed necessary to address remaining uncertainties. However, these confirmatory studies were also non-randomised studies, including follow-up data of the pivotal trials as well as new, non-randomised studies, raising the question whether – beyond the availability of follow-up data – the very uncertainties ascribed to the lack of data from controlled studies could effectively be addressed. In contrast to the EMA, the FDA also granted conditional approval for brentuximab vedotin on the basis of the same non-randomised trials, but required confirmatory evidence from RCTs.[89] Uncertainties regarding the comparative benefits in relation to time-dependent outcomes, including progression-

free survival and overall survival, were therefore only addressed through RCTs imposed by the FDA, but would not have been addressed through the EMA-imposed non-randomised studies in the post-marketing phase (or only addressed partially, with remaining uncertainties associated with comparison of data from single-arm studies to external controls, see section 2.4.2). There are also more recent examples of EMA approvals based on non-randomised studies where confirmatory evidence is expected from further non-randomised studies, including futibatinib for treatment of cholangiocarcinoma (conditionally authorised in 2023 with a requirement to provide confirmatory data from a new study where all patients receive the investigational treatment),[90] selpercatinib for thyroid cancer (conditionally authorised in 2021 with requirement to submit follow-up data from the pivotal, non-randomised study),[91] and tagraxofusp for treatment of blastic plasmacytoid dendritic cell neoplasm (approved under exceptional circumstances in 2021 with requirement for follow-up data from a registry),[92] among others.

As potential sources for pivotal evidence on new medicines, and for addressing uncertainties in the post-marketing setting, non-randomised studies may play a key role in addressing the "evidence vs. access conundrum" and contribute to faster patient access. However, accelerating market access for new medicines through increased use of non-randomised studies while maintaining robust evidence standards relies on some key assumptions which need to be addressed through empirical research.

1.6.1. Internal validity of non-randomised studies and feasibility of RCTs

Methodologically, the key assumption of increased reliance on non-randomised studies is that they can provide internally valid estimates of treatment effects. The use of data from non-randomised studies without very careful consideration of the weaknesses of these study designs "ignores all we know about good research design for identifying causal effects."[93, p1] Non-randomised studies have been shown to be more likely to produce misleading results (i.e. results that were overturned by subsequent studies),[94] and to overestimate treatment effects significantly when compared to RCTs.[95, 96] These issues are particularly relevant for establishing efficacy of new medicines, as overestimation of efficacy may result in a distorted benefit-risk balance. When carefully designed, non-randomised studies may be able to replicate RCT results

for selected clinical questions,[48] but even with the availability of unprecedented levels of details about patients through EHRs and claims data, important confounding factors might not be accounted for, leading to inconsistent and irreproducible findings in observational studies.[97]

Notwithstanding the well-known limitations of non-randomised studies, there is clear interest in using them to establish clinical effectiveness for regulatory purposes. Recent efforts to replicate the findings of RCTs using large, observational data sets showed that some clinical questions (typically for high-prevalence indications) could be addressed through non-randomised studies, although findings for other questions could not be replicated using observational data.[48] It is therefore important to understand whether and under what circumstances non-randomised studies can provide unbiased effect estimates for new medicines.

There are also methodological assumptions why RCTs are not possible or ethical to conduct, and why they may not be necessary in some situations. Potentially valid reasons why RCTs cannot or should not be conducted include instances where not enough patients are available to conduct a RCT in a timely manner, or where it is deemed unethical to randomise patients to a placebo or standard of care arm when there is unmet medical need for that particular indication. Large effect estimates that make it unlikely to be entirely due to bias may also provide reassurance that a non-randomised study has identified a true existing treatment effect. The assumptions underlying an increased reliance on non-randomised studies due to lack of feasibility or lack of necessity for RCTs can be empirically tested.

1.6.2. Regulatory tools to manage uncertainty

Accelerating market access through increased reliance on non-randomised studies reflects an acceptance that products enter the market with some uncertainty about their benefits and harms. Notwithstanding potential solutions to challenges around internal validity (see section 1.6.1 above), regulators continue to acknowledge that non-randomised studies are at increased risk of bias, and that the evidence they generate comes with additional uncertainty. However, this uncertainty may be acceptable if it does not outweigh the benefits of early market access for a new product, and if it can be addressed.

Accepting uncertainty therefore requires a regulatory framework for when it is appropriate to do so and how uncertainty can be managed. Expedited approval schemes provide such frameworks by creating specific pathways to accelerate market access, including some pathways with different evidence standards, and others where evidence requirements can be discussed early between developers and regulators. Conditional approval is an expedited pathway that formalises the management of uncertainty stemming from early clinical trial evidence (including in many cases non-randomised phase II trials) since approval is conditional on remaining evidence gaps being addressed after the product has entered the market. However, empirical studies have raised questions over the extent to which conditional approval and other expedited pathways fulfil the assumption of being able to address the dual requirement of accelerating market access while maintaining robust evidence standards. Criticism of expedited pathways relate to the robustness of pivotal evidence for initial approval and whether confirmatory studies are designed and conducted in a way that is qualified to address uncertainties about the clinical benefit, including whether such evidence becomes available in a timely manner.[62, 70-73] The (timely) generation of robust confirmatory evidence is not only important for regulatory purposes but also to allow downstream decision makers (including HTA bodies, payers, and clinicians) make informed decisions about reimbursement and the routine use of the new treatment in practice.

There are therefore questions regarding the assumption that regulatory tools to manage uncertainty can be implemented and enforced so that regulators ensure that the benefits of early access based on evidence with uncertainties outweigh its harms.

1.7. Thesis structure

Efforts to accelerate market access for new medicines rest on key methodological assumptions about the evidence required to demonstrate that new medicines are safe and effective and about the implementation of regulatory tools to manage uncertainty about their therapeutic value (see section 1.6). Medicines regulation provides a particularly instructive field to study how statutory evidence standards are enforced and how they change in response to perceived patient needs and policy priorities since adherence to the standards set by regulatory bodies decides whether a new product is

allowed to enter the market. By either explicitly setting or implicitly accepting different evidence standards, expedited approval pathways can help address uncertainty about the benefits and harms of new medicines, but also create additional uncertainty.

This thesis aims to examine how medicines regulatory bodies manage uncertainty and presents empirical evidence on the acceptance of non-randomised studies, as well as on the foundational assumptions for accelerating market access for new medicines based on non-randomised studies. The remainder of the thesis is structured as follows:

- **Chapter 2** (literature review) sets the scene through a literature review of conceptual and empirical publications on the key topics addressed in this thesis. Key gaps in the literature are identified and the research questions to be addressed in the thesis are formulated.
- **Chapters 3 to 5** present the findings of the three empirical papers of this thesis.
- **Chapter 6** discusses the main findings of the empirical work presented in this thesis and draws policy conclusions.
- **Chapter 7** provides an overall conclusion of the thesis.

2. Literature review

This chapter presents findings from a literature review on key concepts applied in regulatory science for understanding uncertainty in regulatory approval of new medicines, including study design characteristics and their relevance for expedited and regular approval pathways across the product life cycle, and how regulatory acceptance of uncertainty introduced through different study designs has been studied empirically. The chapter also identifies important research gaps which are addressed through the empirical chapters of this thesis.

2.1. Aims and scope of the literature review

The aims of this literature review were two-fold:

- to understand the role of non-randomised studies and associated study design and analysis characteristics (specifically, the use of surrogate endpoints and different methods for analysing data from non-randomised studies) in contributing to and managing uncertainty in regulatory approval, including in expedited approval pathways, and
- to map existing empirical evidence on the use of non-randomised studies and associated study design and analysis characteristics for regulatory approval.

The literature review first provides a brief overview of theoretical perspectives that may explain developments in regulatory practice in relation to acceptance of uncertainty. This is followed by a detailed review of the sources of uncertainty about clinical benefit of new medicines, i.e. different types of non-randomised studies and their methodological challenges, and how they are used for regulatory approval, including in expedited approval pathways.

For each of these key topics of the thesis, both empirical literature and regulatory guidance documents were reviewed. The sub-sections below primarily focus on empirical literature and make reference to regulatory guidance where relevant. A more detailed review of regulatory guidance documents in relation to different non-randomised study designs and data sources is available in Appendix 1. The review

included both empirical studies and the relevant regulatory guidance and frameworks as complementary sources to understand how evidence standards have been empirically studied and defined in regulatory science. Regulatory guidance describes the (stated) intended role of different types of evidence and provides insights into what regulators deem acceptable in different situations. This may be expected to be followed closely by medicines developers and therefore provides an essential component of understanding regulatory decision making. Empirical studies, on the other hand, provide evidence on actual regulatory decisions, which may or may not be consistent with published guidance.

The literature review focuses primarily (although not exclusively) on cancer since expedited pathways and regulatory acceptance of more uncertain evidence are particularly prevalent in this therapeutic area. Furthermore, the review centres on Europe and the US, whose regulatory bodies – EMA and FDA, respectively – are global leaders in setting evidence standards and developing frameworks for new types of evidence for regulatory approval.[26]

2.2. Theoretical background on developments in regulatory practice

As described in chapter 1, medicines regulators face challenges in balancing different objectives, i.e. to protect and improve public health by ensuring only safe and effective medicines enter the market and to enable access to treatments for patients in need. The introduction of expedited approval pathways (see section 1.4) and a potentially general greater willingness to accept uncertainty (see section 1.3) can be considered responses to this conundrum. They also reflect a shift by regulatory bodies away from their early role as gatekeepers, which focused primarily on protecting public health, towards a position as enablers of pharmaceutical innovation and patient access. Different theories have been proposed to explain this development. While this thesis does not aim to investigate the reasons behind this trajectory, these theories provide a useful background for the empirical analyses conducted in this thesis. Key theories from the literature are therefore briefly summarised below.

2.2.1. Evolving role of regulatory bodies

The theories presented below reflect an understanding that medicines regulatory bodies have become more accepting of uncertainty over time. This understanding is supported by documented legislative changes in both the US and (subsequently) in Europe introducing new, expedited approval pathways that aim to accelerate market access for new medicines, including in some cases explicitly at the expense of more robust evidence at the time of initial approval (see section 1.4 for a brief description of expedited pathways as well as section 2.5.1 for a review of evidence standards in regular and expedited approvals). Even outside expedited approval pathways, regulators are afforded some flexibility in what constitutes acceptable uncertainty. An example is the flexibility for the FDA to determine that a single trial (rather than the prior long-standing regulatory practice of two independent trials) is sufficient to establish effectiveness, which was enshrined in legislation by the US Congress in 1997.[98]

Empirically, an overall higher level of regulatory acceptance of uncertainty has been demonstrated for the US: over a 22-year period, the proportion of approvals of new medicines with randomised trials, trials measuring clinical outcomes, and with more than one pivotal trial has decreased while the proportion of approvals based on studies without control group and measuring surrogate endpoints increased (note that equivalent evidence mapping regulatory trends over several decades is not yet feasible for the EMA, which was only founded in 1995 and initially had a limited scope for approvals of new medicines in selected areas).[36]

Further evidence on the evolving role of regulatory agencies can be found in their mission statements and strategy documents by both EMA and FDA which emphasise their role as facilitators of innovation, partnering with the pharmaceutical industry to enable market access for new, potentially innovative treatments (see section 1.3).[57, 58] The rise of "real-world data" and a product life cycle evidence generation approach, where evidence is generated both before and after market entry, has also been embraced by FDA and EMA, indicating that uncertainty at the time of initial approval may be more acceptable because it is assumed to be addressed in the post-marketing phase.[63-65]

2.2.2. Key frameworks for explaining the evolving role of regulatory bodies

Key frameworks for explaining the evolving role of regulatory bodies have been summarised by social science scholars focusing on pharmaceutical regulation.[65, 99-102] Given the complex interplay between regulatory frameworks, scientific standards, and patient and societal needs, it may not be possible to explain regulatory outcomes through a single theory. Accordingly, different theories are presented below and their potential relevance for medicines regulatory bodies and how they accept or manage uncertainty are briefly discussed.

Public interest theory of regulation

These theories posit that regulation (including, but not limited to medicines) exists to protect population welfare and that regulators act in the public interest.[99] Regulation is considered necessary to correct market failures like imperfect competition, information asymmetry, or externalities. Mission statements of medicines regulatory bodies reflect the public interest (i.e. to protect and improve public health). This theory therefore appears primarily aligned with a more traditional understanding of the role of medicines regulatory bodies. While public interest theory does not explicitly advocate accepting uncertainty in regulatory decisions, additional uncertainty in medicines regulation may be justified from this perspective in specific situations where the trade-off between uncertainty and patient benefit is positive (e.g. in situations with high unmet need).

Regulatory capture

Regulatory capture occurs when industries or sectors exercise influence over the regulatory bodies overseeing them, resulting in regulatory practices favouring commercial – rather than public – interests. Close ties between regulators and the regulated industry may lead to industry values dominating regulatory frameworks.[65, 99, 102] From a regulatory capture perspective, the acceptance of uncertainty about benefit of new medicines – such as approval based on limited clinical evidence – can reflect industry interests in shortening the time to market access for new products and therefore expanding the monopoly period (see also section 1.5 for a description of the phases of evidence generation and their relationship to protection from competition). The empirical evidence on regulatory capture occurring at the FDA is limited, since

other factors, including the FDA's efforts to protect its reputation and political pressure, may better explain its actions and development as regulatory agency.[103, 104]

Corporate bias

Similarly to regulatory capture, corporate bias theory posits that regulators may act in the interests of the industry they are regulating rather than in the interests of the public. However, corporate bias theory goes beyond the immediate relationship between industry and regulators and considers the broader societal context in which these actors operate: regulators may end up acting in favour of private companies because of the industry's privileged access to the state.[99, 100, 105] According to this theory, the systematic prioritisation of industry interests is not necessarily linked to lobbying or direct influence by industry on regulators and law-makers, but through the central role the industry occupies in the state. This theory provides a relevant lens for understanding the shift in the role of medicines regulatory bodies from gatekeepers to enablers of pharmaceutical innovation.

Disease-politics and reputational theories

The needs of patients and their voices play a role in both disease-politics and reputational theories. Through these lenses, regulators are considered to be acting in response to activism of patient advocacy groups and their political power.[99, 100] Reputational theories were famously at the core of Carpenter's analysis of the FDA, which emphasised how reputation as a core source of the agency's authority shapes its actions.[103] Regulatory acceptance of uncertainty in the US has been empirically linked to these theories, as patient activism was considered key for approving new treatments for AIDS through early versions of conditional approval.[105]

Neoliberalism and consumer choice

Finally, regulatory practices and developments can be considered through the lens of neoliberal logic and consumer choice. This perspective assumes that neoliberalism and a focus on the free market (including in health care) shape regulatory actions. In neoliberal logic, medicines are treated as commodity and individuals are framed as consumers rather than patients, which may stretch to pharmaceutical development and the conduct of clinical trials.[106] This perspective may be instructive for comparing regulatory frameworks across regions, in particular comparing the US (with a stronger focus on neoliberalism) and Europe.[99]

The theories presented above present perspectives through which to understand how pharmaceutical regulation has evolved over time. This thesis does not aim to test these theories or to provide evidence on possible reasons for how regulators manage uncertainty. Instead, the thesis examines the methodological assumptions underpinning the observed trend of changing evidence standards and increasing regulatory acceptance of uncertainty. The rest of the literature review therefore focuses on specific design and analysis features of the clinical evidence to support approval of new medicines, focusing on non-randomised studies and how they contribute to uncertainty or the management of uncertainty.

While this thesis does not aim to analyse the structural reasons and drivers behind increasing interest in the use of non-randomised studies, it critically investigates their growing role for regulatory approval and methodological challenges associated with this trend. Focusing on methodological questions, the thesis analyses how evidence standards and the flexibility afforded to regulators are operationalised in practice. In doing so, the thesis contributes to a critical understanding of the role of regulatory bodies and their positioning as protectors of public health vs. enablers of pharmaceutical innovation.

2.3. Study design and its impact on uncertainty

As described in section 1.2.2, there is increased interest in the use of non-randomised studies for regulatory decision making (empirical evidence demonstrating this trend is presented in a separate section of the literature review below, section 2.4). From a methodological perspective, an increasingly important role for non-randomised studies as basis for regulatory approval of medicines may result in additional uncertainty due to their methodological limitations. There are important questions regarding the internal validity of non-randomised studies, which are discussed in more detail in section 2.3.1 below. In addition, uncertainty may be introduced through the relationship between study design and outcome measures: non-randomised studies without concurrent control group – i.e. single-arm trials that lack a control group or are compared to external controls (see section 2.4 below for a description of relevant non-randomised study designs for regulatory decisions) – cannot directly determine the causal effect of a new medicine on patient-relevant outcomes like overall survival or

quality of life; causal inference from these study designs is limited to surrogate endpoints. Uncertainty about clinical benefit of new medicines may therefore stem both from concerns about the internal validity of effect estimates obtained from non-randomised studies and from the translation of surrogate endpoints into meaningful patient-relevant outcomes.

In the sub-sections below, empirical studies on methodological issues with the validity of non-randomised studies and for studies measuring surrogate endpoints are summarised. The literature review includes methodological research on the validity of treatment effects obtained from non-randomised studies and on validation of surrogate endpoints, as well as empirical evidence on how these study characteristics feature in regulatory approvals. Key findings from the literature review are summarised in Box 1.

Box 1: Summary of literature review on study design and uncertainty

What are key methodological issues in relation to non-randomised studies and the use of surrogate endpoints?

- Non-randomised studies introduce additional uncertainty about the existence and magnitude of treatment effects because they are subject to increased risk of bias, including biases that are preventable through randomisation.
- Adherence to high standards for planning, conducting, and reporting non-randomised studies may address risk of bias. However, implementation of advanced methods, such as the target trial emulation approach, may be limited due to data constraints.
- Non-randomised studies without concurrent control group may support causal inference for surrogate endpoints, but do not allow causal inference for patient-relevant endpoints such as overall survival or quality of life.
- Surrogate endpoints may contribute to accelerated patient access to effective treatments when they are validated as predictors of clinically meaningful benefit, but validation is often missing.

How have non-randomised studies and studies measuring surrogate endpoints been used in regulatory practice?

- Non-randomised studies for demonstrating clinical benefits of new medicines often have not applied methods for matching participants across study groups or for adjusting for potential confounders.
- Surrogate endpoints are more likely to be used in expedited approval pathways, but their use is also increasing for regular approvals. The majority of surrogate endpoints used for regulatory approval have not been validated.

- For the majority of new medicines approved based on non-randomised studies, no confirmatory RCT evidence is required. Clinical benefit based on overall survival or quality of life has only been demonstrated for a minority of medicines approved without robust evidence at time of initial authorisation.

Research gaps

- Existing studies comparing effect estimates from randomised and non-randomised studies have not come to a clear conclusion whether and to what extent non-randomised studies provide systematically different effect estimates than RCTs.
- It is unclear whether large treatment effects in non-randomised studies are predictive of clinical benefit of new medicines as demonstrated through RCTs.

2.3.1. Uncertainty over effect estimates from randomised vs. non-randomised studies

When planned, conducted, analysed and reported appropriately, RCTs provide unbiased causal estimates of treatment effects and have therefore long been the mainstay of regulatory approval for new medicines. Effect estimates obtained from non-randomised studies, on the other hand, are subject to increased risk of bias, including biases that are preventable through randomisation.[15, 17] These biases introduce uncertainty about the existence and magnitude of treatment effects. Previous research has examined the internal validity of non-randomised studies in comparison to RCTs and key findings are summarised below.

Meta-epidemiological research

Aiming to better understand these uncertainties, a strand of literature on the internal validity of effect estimates from non-randomised studies has emerged. Given the high internal validity of RCTs, this literature compares effect estimates obtained from the two study types in so-called meta-epidemiological research. Meta-epidemiological research analyses the association of study characteristics with effect estimates.[107] This method has been used to investigate design characteristics of RCTs, such as blinding,[108] but also the absence of random allocation of participants to treatment and control groups.

Meta-epidemiological studies on the comparability of effect estimates from randomised and non-randomised studies in different clinical areas have arrived at different conclusions.[109] In the largest meta-epidemiological study on this topic

(prior to the publication of work presented in this thesis in Chapter 3), published in 2001, Ioannidis et al. found that non-randomised studies were more likely to show larger treatment effects than RCTs.[95] Other early meta-epidemiological studies did not find a systematically exaggerated treatment effect in non-randomised studies.[110, 111] A Cochrane umbrella review found overall no clear difference in treatment effects between the two types of studies.[112] However, there was heterogeneity and some included studies suggested that results differed between RCTs and observational for a subset of between 20% and 40% of outcomes.[95, 113-115]

Some meta-epidemiological studies comparing treatment effects in randomised vs. non-randomised studies have faced criticism for not comparing like with like. Systematic differences between the population in a RCT and the patient population eligible for the treatment of interest can explain why treatment effects obtained from RCTs differ from those seen in clinical practice.[116] Trial participants, often younger and without comorbidities, are not representative of the patient population in clinical practice,[117] and levels of treatment adherence in a trial by both patients and health care providers might not be replicated in a real-world setting. "Real-world" studies aiming to confirm the effectiveness of previous RCTs in clinical practice might therefore be expected to obtain different results. At the same time, non-randomised studies are not immune to poor generalisability to the target population. Moreover, the risk of bias to the internal validity in non-randomised studies may exacerbate issues of poor representation since it introduces additional uncertainty about the direction and extent of potential bias of the effect estimate with respect to the target population.[23]

The methodological debate about the internal validity of treatment effects obtained from non-randomised studies has attracted renewed attention through interest from regulators in the use of real-world (i.e. observational) data. Regulatory bodies have developed frameworks for the use of real-world data in their assessments,[35, 39] and researchers have studied what types of real-world data and non-randomised studies inform their decisions.[54, 118-124] This increased interest has also led to new methodological research on the validity of effect estimates from non-randomised studies. A meta-epidemiological study focused on matched pairs of non-randomised and randomised studies for the same intervention in comparable populations where the non-randomised study was conducted prior to the RCT in order

to overcome some of the criticism faced by previous studies.[96] Using this design, the authors found that studies using observational data overestimated treatment effects by 31% when compared to RCTs, lending further credibility to concerns over biased (exaggerated) treatment effects in non-randomised studies.

There are also other reasons relating to study design that could explain why randomised and non-randomised studies generate different results. Since the two study designs may be used to answer slightly different research questions (e.g. using different definitions for outcomes and the counterfactual), their answers may be internally valid for the specified question. The target trial emulation framework aims to explicitly incorporate these study design elements into the assessment of potential bias in non-randomised studies (see below).[49] While there may therefore be reasons for expecting effect estimates from randomised and non-randomised studies to differ, the availability of evidence from different study types for particular decisions should be considered, such as when granting regulatory approval for a new medicine, or when selecting a treatment for a patient. When evidence from both a RCT and a non-randomised study are available, decision makers may consider the specific question (including definitions of population, outcome, and counterfactual) addressed by each study to inform their actions, potentially leveraging heterogeneity in study designs to arrive at conclusions about the benefits and harms of a medicine for a given population. When this is not the case (e.g. only a non-randomised study exists), the available evidence will still be taken into consideration to address the decision makers' information needs. It is therefore important to understand whether there are meaningful discrepancies about the magnitude and direction of effect between randomised and non-randomised studies.

Overall, meta-epidemiological studies have not demonstrated a systematic over- or underestimation of effect estimates from non-randomised studies compared to RCTs. They did, however, identify important discrepancies for a subset of clinical questions. The lack of a statistically significant systematic over- or underestimation of treatment effects does not mean that RCTs and non-randomised studies generally agree with each other. Across clinical questions, such effects may cancel each other out, but for individual questions, important uncertainties remain.[109] These are important findings for decision makers, including regulatory bodies, because they highlight

uncertainty about both direction and magnitude of treatment effect obtained from non-randomised studies.

Target trial emulation

A different approach to understanding and addressing potential bias in non-randomised studies are target trial emulations. Target trial emulation is a framework for addressing bias in a non-randomised study by explicitly designing it to resemble (or emulate) a RCTs for the same clinical question.[49] This involves specifying a hypothetical target trial (or alternatively identifying a real trial), including defining the causal question to answered, eligibility criteria, intervention details and treatment allocation, time for start of follow-up, outcomes, and the analytical approach including the estimand to be calculated. Once the target trial is defined, replication of all the defined elements is attempted using observational data, such as claims data or EHRs. By addressing common biases in non-randomised studies, the target trial emulation framework strengthens causal inference from observational data.

Regulators are interested in using the framework. An important initiative to demonstrate the feasibility of using target trial emulations for regulatory purposes was RCT-DUPLICATE.[125] Partnering with FDA officials, the initiative aimed to emulate 32 RCTs under the best possible circumstances. For these 32 selected clinical topics, the demonstration project reported a correlation coefficient between RCTs and observational studies of 0.82 (indicating high agreement), full statistical agreement (i.e. point estimates and 95% confidence intervals of both RCT and observational study falling on the same side of the null) for 56% of trials, estimate agreement (i.e. the point estimate of the observational study falling within the 95% confidence interval of the RCT) for 66%, and 75% standardised difference agreement (i.e. the standardised difference between point estimates from RCTs and observational studies).[48]

While the target trial emulation approach is attracting interest as a robust framework for conducting non-randomised studies, its role is still evolving. As observational studies using this design are increasingly published, it is not clear whether the framework is consistently applied in a way that will contribute to less biased estimates. A systematic review of 200 observational studies emulating a target trial from 2012 to 2022 reported severe methodological shortcomings, including failure to describe a protocol for 57%.[126] Another systematic review also identified

methodological issues in how 100 studies were conducted, including failure to specify the target trial in 24% and serious risk of bias in 28%.[127]

Data requirements for implementing a narrowly defined protocol (including rich data on participants, covariates, and outcomes) may further limit the scope of clinical questions that can be addressed in the target trial emulation framework. A review of data requirements for replicating pivotal trials for market approvals of medicines in the US from 2017 to 2019 found that only one of 172 trials could feasibly be replicated using data elements and codes from contemporary real-world data sources.[128]

Uncertainty in regulatory decisions due to non-randomised study designs

The variation in discrepancies between effect estimates from randomised and non-randomised studies creates a major challenge for regulators and other decision makers as it introduces additional uncertainty. While known bias in a systematic direction and magnitude can be taken into consideration, it is more difficult to address uncertainty over the existence and extent of systematic differences in effect estimates obtained from randomised vs. non-randomised studies. However, regulators may be reassured if non-randomised studies are conducted, analysed, and reported appropriately and according to regulatory guidance,[23, 52, 129, 130] and if they show convincingly large effect estimates indicating a true treatment effect,[19, 131] particularly if this is later confirmed through robust evidence. Empirical studies on regulatory approvals have shed some light on how uncertainties related to study design are addressed.

While methodological advances have been made in relation to the planning, conduct, and analysis of non-randomised studies to address potential bias, key features are still often lacking when used for regulatory decisions. Several empirical studies analysing the use of real-world data (typically as sources for external controls) in regulatory decisions have found that the majority did not apply methods for matching participants in the pivotal single-arm trial and the external control or for adjusting for potential confounders (see section 2.4.2 for more details on these studies).[132-134] These shortcomings were also criticised by regulators in their feedback to medicines developers.[54, 122, 132, 134] Regulators also criticised lack of pre-planning (e.g. lack of a pre-specified protocol), lack of comparability between study endpoints, and handling of missing data.

In addition to the existence of a treatment effect, its magnitude can be important for regulatory approval. Larger treatment effects can provide a justification for accelerating market access and not requesting more robust study designs. “Dramatic” effect sizes may provide reassurance that treatment effects observed in non-randomised studies are not entirely due to bias and indicate the existence of a true effect.[131] Indeed, larger effect sizes obtained from non-randomised studies may predict whether regulators grant approval, as shown in a systematic review of the global rate of regulatory approval for cancer medicines.[135]. In that study, the objective response rate (i.e. an endpoint that could provide causal evidence of a treatment effect in a single-arm trial) was significantly associated with whether a treatment had received regulatory approval anywhere in the world. A threshold of 30% response rate was highly likely to predict regulatory approval. The authors interpreted this finding as indicative of a high response rate being an appropriate endpoint for regulatory approval; however, the study did not analyse whether the response rates translated into clinically meaningful benefits.

Investigating treatment effect sizes of non-randomised studies in more detail, Djulbegovic and colleagues reviewed marketing authorisations by the EMA from 1995 to 2015.[136] They found that 51 medicines were approved on the basis of non-randomised studies, corresponding to 7% of all approvals over that period. Using three different thresholds for defining “dramatic” effect sizes, between 29% and 52% of approvals based on non-randomised studies had demonstrated such large treatment effects, corresponding to between 2% and 4% of all approvals over that period. Treatment effects were considerably larger for products for which the EMA did not request subsequent RCTs to be conducted compared to products for which subsequent evidence was requested, indicating that large effect estimates did indeed provide reassurance to regulators. However, the EMA only requested a confirmatory RCT for 28% of the indications. For the majority of approvals based on non-randomised studies, RCT evidence confirming the existence and magnitude of a treatment effect may therefore never materialise. Similarly to Europe, a study of products with breakthrough therapy designations on the basis of non-randomised studies in the US found that only a small proportion of all applications in this pathway fulfilled different thresholds for

large effect sizes, and that larger effect sizes in non-randomised studies were associated with a lower probability of requirements for post-marketing RCTs.[137]

Given the documented methodological limitations in how results from single-arm trials are compared to external controls and lack of confirmatory RCT evidence for many indications, uncertainty over the existence and magnitude of treatment effects may persist in regulatory practice.

2.3.2. Uncertainty in relation to study endpoints

Another key aspect that can introduce uncertainty about the existence of a meaningful benefit is related to the endpoint being studied. In order to understand benefit, outcomes that are meaningful to patients should be measured. These typically include overall survival and quality of life in various domains, including symptoms and pain.[138] The importance of endpoint selection in line with patient preferences is also reflected in internationally harmonised regulatory guidance, which states that "[t]he choice of endpoints should be meaningful for the intended population and may also take into account the views of patients."[17, p17] However, measuring some of these meaningful endpoints can take a long time and/or a large sample size to accumulate the required number of events for a robust statistical analysis. To shorten trial duration, sponsors may rely on surrogate endpoints such as biomarkers and laboratory values instead. There is some empirical evidence showing that surrogate endpoints do indeed reduce the time to approval. In cancer, pivotal trials for FDA approved drugs were shorter when measuring progression-free survival or response rate, saving 11 to 19 months compared to trials measuring overall survival after adjusting for baseline characteristics.[139]

Validity of surrogate endpoints

When validated as predictors of clinically meaningful benefit, surrogate endpoints can therefore contribute to earlier market approval.[140] However, surrogate endpoints may overestimate treatment effects and therefore introduce uncertainty over how beneficial a new treatment is. A meta-epidemiological study found that treatment effects were statistically significantly larger for surrogate endpoints compared to clinical outcomes.[141]

Many surrogate endpoints are not robustly predicting clinical benefit.[142-147] Several frameworks for defining validated endpoints exist, generally resting on evidence on both mechanistic or biological plausibility of predicting clinical outcomes and on statistical validation, such as correlation between surrogate endpoints and clinical outcomes in RCTs.[148] Depending on the type and strength of evidence available, different levels of validation and their robustness can be defined, including biological plausibility, consistent association between surrogate endpoint and clinical outcome, and statistically validated correlation between the two, ideally based on several RCTs.[140]

Endpoint selection is directly linked to the study design: patient-relevant endpoints, such as overall survival and quality of life, require a counterfactual from a concurrent control to provide meaningful evidence. Measuring these outcomes in a non-comparative study is possible but they generally cannot be interpreted meaningfully, as changes may occur in the absence of an effective treatment due to the natural course of disease or the passage of time.[23] Regulatory acceptance of a single-arm trial as evidence on a medicine's effectiveness therefore typically also means acceptance of a surrogate endpoint.

Use of surrogate endpoints for regulatory approval

The use of surrogate endpoints is generally linked to expedited approval pathways as surrogates allow earlier identification of potentially favourable treatment effects. Indeed, surrogate endpoints are more likely to be used in expedited approval pathways, but they also feature in regular approvals. In the US, a comprehensive review of all pivotal trials supporting FDA approvals from 2005 to 2012 reported that surrogate endpoints were significantly more likely to be used in accelerated approvals than in regular approvals.[37] However, the study also showed that surrogate endpoints are common overall, with almost half (49%) of all pivotal trials reporting surrogate endpoints. Clinical trials for cancer medicines approvals with orphan drug designation were also shown to be more likely to be measuring surrogate endpoints compared to approvals for non-orphan drugs.[149] Another study of pivotal trials for FDA approvals found that the proportion of pivotal trials using clinical endpoints decreased from 1995 to 2017 and the use of surrogate endpoints significantly increased during that period overall. While the proportions remained stable for expedited approvals, the trend

towards increased reliance on surrogate measures was driven by approvals without expedited pathways.[36] Within cancer, the number of approvals based on surrogate endpoints has increased over time for both accelerated and regular approvals.[150]

In Europe, surrogate endpoints have been shown to be the basis for the vast majority of expedited approvals.[144] Their use as primary endpoints for EMA marketing authorisations in oncology (including both standard and conditional approvals) has increased over time. While the proportion of marketing authorisation applications based on overall survival decreased by 13% from 2009 to 2017, the use of progression-free survival and overall response rate increased by 13% and 14%, respectively.[151]

Empirical studies have investigated the validity of surrogate endpoints used for regulatory approval and generally reported low levels of validation. For Europe, Schuster Bruce et al. reviewed surrogate endpoints used for products with conditional marketing authorisation or with accelerated assessment designation from 2011 to 2018 and found that the majority of endpoints were only "reasonably likely" (61%) or to have biological plausibility (94%) to predict clinical benefit, i.e. most endpoints were not validated as reliable surrogates for clinical outcomes.[144] In the US, robust evaluations of the validity of surrogate endpoints were only available for 35% of cancer medicine indications approved on the basis of a surrogate endpoint from 2009 to 2014.[152] That proportion was considerably lower for accelerated approvals (16%) than for regular approvals (37%), but for both pathways, the strength of association between surrogate endpoint and clinical outcome was generally low. For some surrogate endpoints, the FDA has published their own validation analysis, but out of 15 validation studies for surrogate endpoints in oncology, only one demonstrated strong correlation with overall survival.[147]

Lack of validation can introduce substantial uncertainty as surrogate endpoints may overestimate clinical benefits.[141] This may be particularly relevant for early trials. In a study of new medicines approved by the FDA on the basis of surrogate endpoints between 2005 and 2012, Wallach et al. compared the effect sizes of the pivotal trials to effect sizes for the same surrogate endpoints used in post-approval studies.[153] They found overall no significant systematic difference in effect estimates between pivotal trials and post-approval studies. However, for a substantial majority (75% of trials with

non-continuous surrogate endpoints and 55% of trials with continuous surrogate endpoints), the pivotal trials showed larger effect estimates than the post-approval studies.

Other studies have investigated whether treatment effects observed for surrogate endpoints translated into clinical benefit. In a review of all cancer medicines with FDA accelerated approval from 2013 to 2023 (i.e. approvals based on surrogate endpoints that were "reasonably likely" to predict clinical benefit at the time of approval), less than half had demonstrated a benefit in overall survival or quality of life in confirmatory studies within five years after approval.[154] Similar findings were reported for accelerated approvals of cancer drug indications, for which only 20% to 21% had demonstrated overall survival benefit in post-marketing trials.[142, 155]

The literature shows that surrogate endpoints are an important study design feature to help accelerate market access for new treatments, as they occur quicker compared to clinical benefits, such as overall survival, and can increase statistical power to detect pre-specified treatment effect sizes. However, they introduce uncertainty about the clinical benefit when not validated across several studies. Treatment effects for surrogate endpoints may not translate into the same treatment effect on clinical outcomes. Without confirmatory evidence, the uncertainty about clinical benefit may persist after a medicine is available on the market.

2.4. Role of non-randomised studies for regulatory approval

Non-randomised studies have long featured in regulatory decision making. For example, observational studies can provide evidence on disease prevalence or incidence and the natural history of disease. Observational studies have also long played an important role in pharmacovigilance. In recent years, non-randomised studies have attracted renewed interest as sources for determining the clinical benefit of new medicines for regulatory approval, including as pivotal evidence for initial approval and for supplemental indications, and as confirmatory evidence in the post-marketing setting. The latter is common in the context of conditional approvals: In the first ten years of the conditional marketing authorisation pathway in the EU, 42% of post-marketing studies to address uncertainties at the time of conditional approval were non-randomised studies.[156]

The focus of this thesis is on the use of non-randomised studies for establishing clinical benefit, therefore only study designs relevant for this use are discussed below. The literature review focuses on two interventional non-randomised study designs (i.e. single-arm trials and trials with external controls) and one non-interventional design (i.e. observational cohort studies), and also includes real-world data as source for non-randomised studies.

2.4.1. Single-arm trials

Single-arm trials are interventional studies where all participants receive the same treatment and are followed prospectively. This study design does not include a control group, although data from a single-arm trial could be compared to data from external controls (see section 2.4.2). From a methodological perspective, this study type also includes interventional studies with multiple arms but where no formal comparison is made between these arms. While technically containing more than one arm, each of the arms in such a study would be considered a single-arm trial.[157]

In the following sub-sections, methodological challenges of single-arm trials are described, followed by a brief review of empirical studies on how single-arm trials have been used for regulatory approval. Box 2 summarises key findings of the literature review.

Box 2: Summary of literature review on single-arm trials

What are key methodological features of single-arm trials (SATs)?

- Lack of control group means that a causal effect can only be demonstrated in a SAT for an endpoint that would not occur without a biologically plausible mechanism, such as tumour response.

What are key elements in regulatory guidance on SATs?

- Studies with concurrent control group, in particular RCTs, are generally preferred for demonstrating efficacy, even when they are considered challenging to conduct. SATs may be acceptable when RCTs are not feasible.
- Measures should be taken to mitigate the increased risk of bias in SATs.

How have SATs been used for regulatory approval?

- The proportion of pivotal trials without control group for regulatory approval has increased over time, and regulatory acceptance of SATs overall was shown to be high, with few applications refused authorisation.
- SATs have overall more commonly been used in expedited approval pathways, but evidence from the US also shows use for one third of regular approvals in cancer.
- The majority of new cancer medicines approvals are now based on SATs in both Europe and the US.

Research gaps

- Given their methodological challenges, regulatory guidance suggests a limited role for SATs under specific circumstances only. While approvals based on SATs are now common in some therapeutic areas, existing studies have not empirically investigated regulatory acceptance of SATs.
- Moreover, lack of feasibility of RCTs in cases where approval was granted based on SAT evidence has not been analysed extensively.
- There is a lack of studies to comprehensively investigate the emergence of robust evidence (i.e. RCTs) on clinical benefit after approval based on SATs.

Abbreviations: SAT, single-arm trial; RCT, randomised controlled trial.

Methodological challenges in single-arm trials

Due to the lack of control group, a causal relationship between the studied intervention and the outcome can only be established for endpoints that would not occur for a single enrolled participant without a biologically plausible effect of the treatment being investigated.[23] Accordingly, the selection of endpoint is of paramount importance when designing a single-arm trial. In addition to the absence of a control group, single-arm trials also lack other key features that mitigate risk of bias

in comparison to the commonly accepted gold standard for regulatory approval, double-blind RCTs: random allocation to treatment, enrolment of participants without knowledge about the treatment they will receive, and blinding of participants, investigators, and outcome assessors.[13, 23] An empirical analysis of single-arm trials used for FDA accelerated approvals of oncology medicines from 1992 to 2020 showed that all trials were at high risk of bias and ascribed the observed effect estimates in these trials to possible bias.[158] Methodological challenges in determining causal effects from single-arm trials are reflected in guidance issued by regulatory bodies, resulting in a general stated preference for RCTs.[16, 18, 19, 23] Regulatory guidance issued by EMA and FDA is summarised in Appendix 1 .

Empirical studies on use of single-arm trials

Several empirical studies have analysed the characteristics of pivotal single-arm trials and their acceptance for regulatory approval. These studies have shown that uncontrolled trials are most commonly used in expedited pathways and that they commonly feature as pivotal evidence in oncology but are also used in some cases in other therapeutic areas. While regulatory guidance documents emphasise the role of RCTs as primary source of evidence, they describe some situations where single-arm trials, rather than RCTs, are acceptable. However, no empirical studies have been identified that focused on justifications for accepting single-arm trials.

In the US, the proportion of pivotal trials without a comparator (i.e. uncontrolled studies) has increased and the proportion of pivotal randomised trials has decreased overall from 1995 to 2017. These trends were statistically significant for approvals through expedited pathways, but not for approvals without any expedited programme.[36] Previously, Downing et al. had analysed all pivotal trials for FDA approvals from 2005 to 2012 and reported that pivotal trials were statistically significantly more likely to be non-randomised for approvals with orphan drug designation and for accelerated approvals.[37] While single-arm trials are more commonly used in expedited approval pathways, they can also feature in regular approvals. A study of all cancer medicines approved in the US based on single-arm trials from 2002 to 2021 showed that 66% of approved indications were accelerated approvals and 34% were regular approvals.[28]

Single-arm trials are particularly common as pivotal evidence for cancer medicines. Studies without a comparator accounted for the majority (53%) of pivotal trials in the US from 2005 to 2012 in oncology, but not in other therapeutic areas.[37] Similarly, a study analysing EMA and FDA approvals without any data from RCTs from 1999 to 2014 reported that the majority (66%) of these approvals were in oncology, followed by rare metabolic conditions.[157] In the EU, 55% of solid tumour medicines approved from 2019 to 2021 were based on pivotal evidence from single-arm trials, up from 15% between 2012 to 2018.[29]

Within oncology, a comprehensive analysis of all approvals of cancer medicines in the US based on single-arm trials (including both accelerated and regular approvals, and approvals for first and supplemental indications) characterised the indications for which these trials were used. The study showed that the vast majority of approvals (99%) were for treatment of locally advanced or metastatic disease, and 40% were for treatment of a disease that was defined by the presence of a biomarker.[28] Single-arm trials therefore appear most relevant in more advanced treatment settings, but they are not restricted to biomarker-defined sub-populations.

In a study of cancer medicines approved by the EMA based on pivotal single-arm trials from 2010 to 2019, Tenhunen et al. identified 22 such approvals.[31] These were mostly conditional marketing authorisations. The pivotal single-arm trials all used response rates as primary endpoint, and results for objective response rate ranged from 15.8% to 88.0% (median, 55.6%). The thresholds for determining efficacy based on response rates in the trials were typically pre-specified, although these were often lacking a clinical justification. Mulder et al. also investigated the justifications for thresholds for clinically meaningful benefits in single-arm trials of EMA-approved cancer medicines and identified such justification for 10 of 21 trials.[29]

Endpoints of all single-arm trials used for approval of cancer medicines in the US were analysed by Agrawal et al.,[28] showing that 98% used response rate as primary endpoint. Accelerated approvals for cancer medicines based on pivotal single-arm trials from 1992 to 2020 were further characterised by Ribeiro et al.[30] The study documented an increase in the number of accelerated approvals based on single-arm trials since 2010, with a notable uptick from 2017 to 2020 and noted a lack of methods to account for potential confounding in all trials.

Empirical studies have overall focused on the characteristics of pivotal trials and, where these were single-arm trials, mostly analysed their characteristics.[28-30, 133, 157, 159] The reasons for regulatory acceptance of single-arm trials (i.e. the application of regulatory flexibility in evidence standards), on the other hand, have attracted less attention by researchers. While there is common acceptance that RCTs may not be feasible in small patient populations without effective treatment option, the validity of these reasons has not been analysed extensively. In addition to some studies documenting feasibility of RCTs for small patient populations in general,[160, 161] the literature review only identified one study specifically addressing the feasibility of RCTs for medicines approved based on single-arm trial data, and this was limited to a hypothetical analysis for the US: Rittberg et al. determined the sample size, comparator, and time frame required to conduct hypothetical RCTs for cancer medicines approved by the FDA from 2010 to 2019 and reported that such RCTs would have been possible for the vast majority of approved indications.[159]

Given the well-known methodological limitations of single-arm trials, confirmation of clinical benefit may may indicate appropriate regulatory flexibility at the time of approval. In the US, Agrawal et al. assessed whether cancer medicines approved based on single-arm trials through the accelerated approval pathway had fulfilled the requirements for confirming clinical benefit and reported that was the case for 38%, with 52% pending confirmation, and 9% withdrawn from the market.[28] However, the study did not investigate the characteristics of studies used to confirm clinical benefit. Other studies following medicines with accelerated approval (although not focusing on products approved based on single-arm trial evidence) have shown that confirmatory trials often have similar methodological features as the early trials used for initial approval (i.e. they may lack randomisation and may continue to measure surrogate endpoints), raising questions over the robustness of confirmatory evidence.[61, 142] Findings of these studies are described in more detail in section 2.5.2. However, no studies were identified that comprehensively investigate whether clinical benefit of medicines approved through single-arm trials was later confirmed through RCTs. Further, while some evidence exists on confirmation of clinical benefit for medicines approved based on single-arm trials for FDA-approved products, no studies exist on the emergence of confirmatory evidence after EMA approval based on single-arm trials.

2.4.2. Trials with external (or historical) control arms

The lack of a control group in single-arm trials creates substantial uncertainty due to challenges in the clinical interpretation of the trial findings. When used as pivotal evidence supporting the approval of a new medicine, this leaves regulatory bodies to manage the resulting uncertainty. Controls are needed to contextualise the outcomes observed, as illustrated by the following quote drawn from guidance on choice of control group in clinical trials by the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH), an organisation bringing together regulatory bodies (including EMA and FDA, among others) and industry to harmonise regulatory guidance:

> *"Control groups have one major purpose: to allow discrimination of patient outcomes (for example, changes in symptoms, signs, or other morbidity) caused by the test treatment from outcomes caused by other factors, such as the natural progression of the disease, observer or patient expectations, or other treatment. The control group experience tells us what would have happened to patients if they had not received the test treatment or if they had received a different treatment known to be effective."* [162, p3]

Decision makers are therefore interested in comparing the results from single-arm trials to an external control group. The "external" label relates to the fact that participants in the control group were enrolled separately from (rather than concurrently with) the participants in the trial. Common sources for external controls are past clinical trials or observational studies on the natural history of disease in the same or related patient group. When existing data are used as external control arm, they represent historical controls, and the two terms are sometimes used interchangeably. An external control group can also be created from real-world data, such as administrative and claims databases or hospital records. This allows the control group to be defined while the trial is ongoing or after it has ended; the data would therefore not be historical but would still be external to the trial (i.e. collected in a different setting). Different external control groups can also be combined to create a synthetic control group which more closely resembles the participants in the trial.

Some examples of approvals based on single-arm trials with external controls include the initial (conditional) approvals of blinatumomab for treatment of acute lymphoblastic leukaemia (ALL), cemiplimab for treatment of squamous cell carcinoma, and dinutuximab beta for treatment of neuroblastoma, and regular approvals of axicabtagene ciloleucel for treatment of lymphoma, rituximab for treatment of lymphoma and leukaemia, and defibrotide sodium for treatment of hepatic veno-occlusive disease.[39, 122, 133]

In the following sub-sections, methodological challenges of trials with external controls are summarised, followed by a review of empirical studies on the use and methodological features of externally controlled trials for regulatory approval. Box 3 summarises key findings from that literature review.

Box 3: Summary of literature review on trials with external controls

What are key methodological features of trials with external controls?

- Studies with external controls are at increased risk of bias compared to RCTs. Risk of bias arises due to potential confounding, but externally controlled studies also face other sources for risk of bias.

What are key elements in regulatory guidance on external controls?

- Due to their increased risk of bias, externally controlled studies should only be used in specific situations, in particular when there is a dramatic treatment effect and the usual course of the disease is highly predictable.
- Regulatory guidance documents do not specify appropriate analytical methods but recommend adherence to general design and analysis principles for addressing risk of bias.

How have external controls been used for regulatory approval?

- External controls have featured in both expedited and regular approval pathways.
- The majority of studies used for regulatory approval use naïve comparisons, rather than advanced analytical methods.
- Comparisons to external controls are commonly not made explicit.

Research gaps

- While external controls are considered acceptable for regulatory approval only in limited situations, potential justifications for regulatory acceptance, such as dramatic treatment effect, have not been investigated comprehensively.

- There is limited evidence on the internal validity of effect estimates obtained from SATs with external controls, and whether and how this is impacted by the use of different data sources and analytical methods.

Abbreviations: SAT, single-arm trial; RCT, randomised controlled trial.

Methodological challenges related to external controls

While the use of an external control arm approach allows for a comparative analysis of treatment effects, the external nature of the control group increases the risk of bias.

A key risk of bias arises from an uneven distribution of baseline characteristics between participants in the intervention trial and the control group that are predictive of the treatment received (i.e. bias due to confounding).[15] This can distort the true treatment effect if the two groups differ systematically from each in characteristics that affect the outcome. Differences can exist for a wide range of participant characteristics, including some that are not known or not measured. Risk of bias due to confounding was previously commonly referred to as "selection bias", and still features as such in regulatory guidance.[16, 162] However, risk of bias due to confounding is not related to selection of participants into the study; rather, confounding may also occur when the same inclusion and exclusion criteria apply to intervention and control groups.
A distinct risk of bias related to selection arises when participation or inclusion in the study (e.g. through inclusion/exclusion criteria when identifying a cohort of patients) is related to both intervention and outcome.[15] An example of how differential selection into the study can distort effect estimates is through immortal time bias, where participants in the treatment group must have survived in order to receive the treatment (i.e. they experienced a period where they were "immortal" for the study) while participants in the control group have been followed since diagnosis and being at risk of death throughout the study period.

Externally controlled studies may also be at particularly high risk of bias in the measurement of outcomes. This is partly due to lack of blinding to the intervention received, although this risk also exists in open-label RCTs. Externally controlled studies face an additional challenge in ensuring that potential differences in what outcomes are available and how they have been assessed in the clinical trial and the external control

do not affect the result. An example is the use of different criteria for defining objective response in the trial and the external control group.

Other sources of bias include the lack of blinding more generally and the retrospective identification and selection of a (historical) control group.[16, 162] Finally, the clinical environment (including diagnostics and supportive care) of external controls can be substantially different.[52] Even when an external control group comes from the same or similar institution, the standards may have changed over time.

The use of external controls therefore calls for analytical methods to address potential differences between participants in the interventional trial and the control group. Several approaches exist and have been used to estimate treatment effects of health interventions, including adjustment for participant characteristics,[44] methods for matching participants in intervention and control groups,[44, 47, 163] and exploiting natural variation in the use of different interventions (instrumental variables approach).[44, 164] Propensity score methods, in particular, have attracted attention to support regulatory decision making.[44, 48, 165] Propensity scores estimate the probability of receiving a treatment given a participant's characteristics. The propensity score method can then be used as weight in adjusted analysis or to match or stratify participants across study groups. However, even advanced methods like propensity score methods and other analytical approaches may not be able to fully address risk of bias in non-randomised studies.[166]

While several analytical methods exist for addressing potential risk of bias due to confounding, the simplest use of an external control group is a naïve comparison, i.e. comparing results from a trial to an external control group without making any adjustments for potential discrepancies between participants. Indeed, naïve comparisons may be the most common use of external controls.[133, 167]

Recognising the methodological limitations of external controls, regulatory guidance by the EMA and FDA limits their use to selected situations, i.e. when the treatment effect is large, the natural history of disease is well-known, and endpoints and their relationship with covariates are well understood.[16, 162] Appendix 1 provides an overview of regulatory guidance in relation to external controls. [16, 17, 40, 162]

Empirical studies on use of external controls for regulatory approval

While some empirical studies exist on the role of single-arm studies for regulatory approval, only few were identified that focused on the use of external controls. Identified studies mostly focused on cancer indications and showed that single-arm trials with external controls are used in both regular and expedited approval pathways. A common theme of the studies identified in this literature review is that many approvals based on single-arm trials lack explicit external controls, and that when they are specified, they are used in naïve comparisons. Similar to the literature review on single-arm trials (see section 2.4.1), no empirical studies were identified that analysed justifications for relying on trials with external controls, rather than concurrently controlled RCTs, for regulatory approval.

Regulatory decisions by EMA and FDA for applications based on pivotal single-arm trials between 2005 and 2017 and the use of external controls in these were characterised by Goring et al.[133] The study showed that explicit consideration of external controls was not common: among 43 included indications, only four included explicit comparisons to external controls based on individual patient-level data, and another 12 used external controls without individual patient-level data (e.g. summary data from previous studies). The remaining 27 indications (63%) did not include explicit external controls. Among the six sources with individual patient-level data, four used various approaches to match external controls with the single-arm trial, including applying matching eligibility criteria, matching pairs of patients, propensity score weighting and stratification, and weighted stratified analysis.

Goring et al.'s study also revealed overall broad acceptance of single-arm trials, with a 79% approval rate by the EMA (approval rate by the FDA was also reported, but since the FDA does not systematically publish reports of unsuccessful applications, these figures are not considered robust).[133] However, analysis of the seven applications without favourable assessment by the EMA revealed that the quality of external controls played a role in three (i.e. the external controls were not considered reliable), and lack of a comparator was deemed relevant for the refusal to grant approval for another three.

Reviewing regulatory approvals through expedited pathways from 2015 to 2021, Sola-Morales et al. compared how uncontrolled trials used real-world data-derived

external controls in EMA vs. FDA submissions.[134] Such controls were more common in FDA submissions compared to EMA submissions, but overall increased over time. The study identified differences in how external controls with real-world data were used: while the vast majority (82%) of FDA submissions used external controls as descriptive benchmarks, EMA submissions used them both in direct comparative analyses (65%) and as descriptive benchmarks (58%), i.e. without direct comparison. The study also included external controls derived from real-world data for HTA submissions (findings not reported in this literature review).

Focusing on cancer medicines, Wang et al. reported that 18 of 103 (17%) marketing authorisations by the EMA from 2016 to 2021 used external controls.[122] Of these, 39% received conditional marketing authorisation. The study showed that the vast majority of external controls were used to analyse the comparative efficacy of the new treatment or to provide a historical benchmark. Other uses included the definition of the margin of non-inferiority for a single-arm trial and understanding the natural history of disease. The majority (54%) of external controls were used in a naïve comparison while 29% were used in comparative efficacy analysis with adjustment or matching based on individual patient-level data.

Ribeiro et al. reviewed the historical controls used to contextualise treatment effects obtained from single-arm trials for cancer medicines with FDA accelerated approvals from 1992 to 2020.[30] Categorising the description of historical controls or comparator treatments in FDA review documents into fully described (i.e. a clear description of the historical control or the current best treatment option), partially described (i.e. the FDA made some reference to existing treatment options), and not described (i.e. the FDA did not include information on a comparator treatment), the study found that clear and explicit descriptions were only available for one quarter of the products.

As described above, some empirical studies have documented the use (or lack) of analytical methods for comparing external controls to participants in a pivotal trial. These studies did not assess the internal validity of the different methods applied or the potential impact of the use of different sources for external controls (such as real-world data or previous clinical trials) on the internal validity of effect estimates obtained from

these studies. These methodological questions are the focus of another strand of literature which was described in section 2.3.1 of the literature review.

2.4.3. Observational cohort studies

Differently from the study designs described above, observational studies are non-interventional, i.e. they do not include a protocol for providing the treatment under investigation to enrolled participants. For medicines regulatory purposes, a group (cohort) of participants is identified based on the intervention they received and followed over time to observe outcomes of interest. The cohort can be compared to another cohort to provide comparative evidence on an intervention's effects. Cohorts can be defined both prospectively (i.e. identifying participants prior to study initiation and following them over time) and retrospectively (i.e. identifying participants with the relevant exposure based on historical data). In addition to cohort studies, case-control studies are another relevant observational study design for regulatory purposes, although these are more relevant for safety studies than for studies aiming to demonstrate effectiveness.

In the sub-sections below, empirical studies on the use of observational cohort studies for regulatory approval are summarised. Box 4 synthesises key findings.

Box 4: Summary of literature review on observational cohort studies

What are key methodological features of observational cohort studies?

- Observational studies are at increased risk of bias compared to RCTs, including selection bias, information bias, time-related bias, and bias due to confounding.
- Target trial emulation provides a framework for understanding and addressing risk of bias in observational studies.

What are key elements in regulatory guidance on observational studies?

- Guidance from Europe and the US stresses transparency in conducting, analysing, and reporting. Guidance on studies using real-world for regulatory decision-making data is still evolving.

How have observational cohort studies been used for regulatory approval?

- Observational cohort studies have been used as pivotal studies for initial approvals and extensions of existing indications.
- They are most commonly used in the post-marketing phase to generate additional (confirmatory) evidence on efficacy and safety.

Research gaps

- Observational studies may form part of the cumulative evidence on benefits and risks of new medicines, but their role as one of the regulatory tools to manage uncertainty has not been comprehensively studied.
- Open questions remain regarding the internal validity of effect estimates obtained from observational studies.

Methodological challenges in observational studies

Methodological questions regarding the internal validity of effect estimates obtained from observational studies remain. A body of meta-epidemiological studies comparing effect estimates from randomised and non-randomised studies (including observational cohort studies) has emerged over the past decades, with heterogeneous results regarding the internal validity of non-randomised studies. This literature was presented in section 2.3.1 of the literature review above.

Both EMA and FDA have published documents outlining their thinking about planning, conducting, analysing, and reporting observational studies, including the use of real-world evidence for regulatory purposes (see Appendix 1 for a summary of regulatory guidance on observational studies).[39, 49, 52, 53, 129, 130, 168, 169] These documents show the interest of regulators in using such study designs for regulatory

decision-making across the pharmaceutical product life cycle and emphasise transparency in methods and reporting.

Use of observational studies for regulatory approval

Observational studies have been used to generate evidence on benefits and harms of medicines and to support regulatory decisions in both pre- and post-marketing phases.[52]

Cohort studies may be used in the post-marketing setting to generate additional evidence on effectiveness and safety and therefore to address remaining uncertainty (e.g. to confirm clinical benefit in a specific population). While interventional studies are also commonly used to address open questions about clinical benefit, observational studies have been used to confirm effectiveness and safety by both EMA and FDA. In a cohort study of all new medicines with conditional EMA approval from 2006 to 2014, 8% of the required post-marketing studies were observational, and their primary study objective was split evenly between safety and effectiveness.[74] In the US, observational studies were specified as real-world data-based post-marketing obligations to assess long-term safety of cancer medicines.[123] Indeed, observational studies are a key source of evidence on safety in the post-marketing setting, as shown in a study of safety and efficacy referrals to the EMA which found that post-marketing assessments relied on similar levels of safety evidence from non-interventional studies and randomised trials.[170]

However, observational studies have also been used to support approval decisions, including for new medicines (i.e. in the pre-marketing phase) and for extensions of existing approvals (i.e. in the post-marketing phase). For example, an observational cohort study was used to provide evidence on effectiveness and safety of palbociclib to treat breast cancer in male patients.[54] The extension of the indication from female to male patients was approved by the FDA. Pre-marketing, observational cohort studies can act as source for the (external) control group for uncontrolled trials (see also section 2.4.2). For EMA cancer medicines approvals using external controls from 2016 to 2021, observational studies were used as source for external controls in 21%. An example is the conditional approval of belantamab mafodotin for the treatment of multiple myeloma where published observational studies provided the benchmark for the

treatment effect observed in a single-arm trial.[122] Since observational studies use real-world data, additional details on their use for regulatory approval decisions are provided below in section 2.4.4.

A common feature of the studies referenced above and in related sections of the literature review is that they analyse the use of and characteristics of observational studies for regulatory decision making at the aggregate level, i.e. they present figures on the number or proportion of approvals featuring observational studies (or data from observational studies) in some form. While these studies shed light on characteristics of observational studies, there has been less focus on their role as part of a cumulative evidence package. As regulators increasingly adopt a life cycle approach to assessing clinical evidence, observational studies need to be analysed in the context of other evidence available for assessing the benefit-risk balance of new medicines.

2.4.4. Real-world data in regulatory decisions

Closely related to study design are the data sources used in non-randomised studies. Real-world data have attracted increased interest for establishing clinical benefit. Real-world data are defined as "routinely collected data relating to patient health status or the delivery of health care from a variety of sources other than traditional clinical trials (e.g. claims databases, hospital data, electronic health records, registries, Mhealth data, etc.)."[41, p1] The evidence generated from studies using real-world data to assess the use and potential benefits and harms are referred to as real-world evidence.[39]

Real-world data have been used for a range of regulatory use cases, including for understanding disease epidemiology and to inform clinical trial design, as well as monitoring medicines utilisation post-approval and safety.[40] There is increased interest in using real-world data to study a medicine's effectiveness in two further use cases: providing evidence on the benefit-risk ratio of a medicine before approval, and providing additional or confirmatory evidence after approval. As described in sections 2.4.1 through 2.4.3, non-randomised studies to determine clinical efficacy and safety can be interventional or observational, and real-world data can play a role for both of these study types: for interventional single-arm studies, real-world data can act as external

control, while observational studies (i.e. studies that are not clinical trials), by definition, typically use real-world data.

In the sub-sections below, existing literature on the use of real-world data to inform regulatory decisions before approval and in the post-marketing setting are summarised. A brief review of regulatory guidance on the use of real-world data is provided in Appendix 1, identifying existing frameworks and ongoing work relating to real-world data for regulatory decision-making across the product life cycle.[26, 34, 35, 39, 129, 171-174] Box 5 summarises key findings from the literature review on empirical studies and regulatory guidance.

Box 5: Summary of literature review on real-world data

What are key methodological features of real-world data?

- Non-randomised studies using real-world data are at increased risk of bias compared to RCTs.
- Design and analytical methods to mitigate risk of bias exist but are not commonly applied in studies using real-world data for regulatory approval.

What are key elements in regulatory guidance on real-world data?

- Guidance on studies using real-world for regulatory decision making data is still evolving, including on specific topics such as data quality and real-world data in observational studies and clinical trials.
- EMA and FDA documents suggest a particular interest in using real-world data to demonstrate clinical benefit for supplemental indications and in the post-marketing setting.

How have real-world data been used for regulatory approval?

- Real-world data have been used for initial approval and to confirm clinical benefit. Their use has increased over time.
- For initial approvals, real-world data have most commonly been used as external controls for single-arm trials.
- Real-world data may also feature as sources for additional or confirmatory benefit in the post-marketing setting for the majority of new medicines. Most commonly, the focus of post-marketing studies using real-world data has been on safety.

Research gaps

- There are open questions regarding the internal validity of effect estimates obtained from studies using real-world data, and studies with different design and analytical features.
- Studies using real-world data may play a role in addressing uncertainty, including both for initial approval and as confirmatory evidence, but existing studies have not focused on this aspect and instead provided high-level overviews of different applications of real-world data.

Empirical studies on the use of real-world data for regulatory decisions

Empirical studies on real-world data for regulatory approval have documented their use in regulatory decisions, including for initial approvals, where they are mostly used as external controls, and in the post-marketing setting. Common data sources include EHRs and registries. Findings from key empirical studies focusing on the US and Europe (although in some cases also including experiences from other countries)

are summarised below. Notably, these studies have not investigated the internal validity of effect estimates obtained from studies using real-world data.

While real-world data feature in the pre-approval phase (i.e. for marketing authorisations of new medicines), they are more commonly used in the post-marketing setting. In a review of EMA decisions involving real-world data from 2018 to 2019, more than two thirds of marketing authorisation applications featured real-world data in the post-authorisation phase, with the rest split between pre-authorisation phase (i.e. providing evidence prior to approval) and a combination pre- and post-authorisation phases.[175]

However, real-world data feature throughout the product development process, as shown by Eskola et al. for a cohort of all 111 new medicines authorised by the EMA from 2018 to 2019.[176] The study reported that real-world evidence use was mentioned in public regulatory review documents for almost all included products at some point during the development cycle. Such use was identified in the discovery phase for 98% of products, and for the purpose of life cycle management (including safety monitoring) for 100%. Approximately half of products used real-world evidence for clinical development (including trial design and comparative efficacy and safety) and registration (i.e. to demonstrate clinical benefit).

In a separate study, Flynn et al. also characterised how real-world evidence was used in EMA approvals for new medicines and for supplemental indications from 2018 to 2019.[175] The study reported real-world evidence being used in 63 of 158 (40%) applications for new medicines and in 28 of 153 (18%) applications for extensions of indications for existing products. For new medicines using real-world evidence, 15% included this as main (or pivotal) study, 75% as supportive study, and 10% as both main and supportive study. The proportion using real-world evidence as main study was comparatively higher among applications for indication extensions (25%). The primary interest was safety for more than 80% of real-world evidence studies, and approximately half of the studies assessed efficacy. The most common source for real-world data were registries, followed by hospital data.

The relative importance of real-world data for initial approvals vs. approvals of line extensions was assessed differently in a study of regulatory decisions based on real-world data in the EU, US, Canada, and Japan.[118] In this study, only two of 26

identified decisions related to approvals of new indications, and four to extensions of the existing indication to additional subsets of patients. Real-world data sources for these uses included both patient registries and medical records.

Purpura et al. reviewed use of real-world data in FDA approvals for new medicines from 2019 to 2021, reporting that 116 of 136 (85%) eligible approvals included such data in some form.[121] In most cases (65%), the applicant intended to use the data as evidence on effectiveness and/or safety. The FDA used the data as primary evidence in a minority (9%) and more commonly used it as supportive evidence (65%).

Real-world data as source for external controls were systematically reviewed by Sola-Morales et al. for approvals through EMA and FDA expedited pathways.[134] The authors reported a substantial increase in the number of submissions using real-world data at both agencies from 2015 to 2021. The vast majority (89% of EMA submissions and 84% of FDA submissions) were for initial approvals, rather than supplemental indications. For both EMA and FDA, the most commonly used real-world data source were EHRs (46% and 45% of EMA and FDA submissions, respectively). Other commonly used sources included the combination of sources (in particular for the EMA), literature (more common for FDA submissions compared to EMA submissions), and registries (more common for EMA submissions).

A mix of real-world data sources for external controls was also reported in other studies. In a review of 26 regulatory decisions for pharmaceuticals by the FDA, EMA, Health Canada, and the Japanese Pharmaceuticals and Medical Devices Agency using real-world data, the majority (65%) were for initial approvals and the rest for approvals of supplemental indications.[118] For initial approvals, real-world data were primarily used as historical controls, and the vast majority of data came from medical records.

Focusing on cancer, a study of EMA approvals with external controls from 2016 to 2021 reported that 37% of the external controls were derived from real-world data, 29% from historical clinical trials, 4% from both real-world data and clinical trials, and 21% from published observational studies.[122] A systematic review of the use of external controls in regulatory approvals by the EMA and FDA in selected therapeutic areas showed that the majority of applications based on single-arm trials did not include external controls.[133] When individual patient-level data were used as external controls, these were mostly based on real-world data.

Approvals of cancer medicines using real-world data were also studied in the US. Arondekar et al. focused on the use of real-world evidence to support efficacy in initial and supplemental approvals for oncology products from 2015 to 2020.[132] They identified 11 of 133 initial approvals (8.3%) featuring real-world evidence to support efficacy, and two of 573 supplemental approvals (0.8%). For new drug approvals, real-world data were mostly used for contextualisation of treatment effects (i.e. descriptively), although five of the 11 approvals also included statistical comparisons. All real-world evidence studies were retrospective, and the most common data source were medical chart reviews.

The use of real-world data in post-marketing obligations issued by the EMA and FDA was analysed by Mofid et al.[119] Identifying 109 EMA-approved products from 2007 to 2020 and 56 FDA-approved products from 1998 to 2020, the study reported that registries were the most common real-world data source for post-marketing obligations across both agencies. In both markets, the predominant use of real-world data-based post-marketing obligations was to collect safety data, although efficacy was also the focus of a minority of obligations.

Focusing on the US, Zettler reviewed the use of real-world evidence to address post-marketing obligations for cancer medicines.[123] From 2017 to 2020, a total of 456 post-marketing obligations (including post-marketing requirements and post-marketing commitments) were issued by the FDA for both new drug approvals and approvals of supplemental indications. Only 15 of these included an explicit request for real-world evidence, and they were mostly issued for new drug approvals. The largest category of real-world evidence requests in post-marketing obligations related to safety reports (47% of obligations), followed by prospective observational studies (27%), and data from expanded access studies and existing registries (20%, respectively).

Similar to the review on observational cohort studies presented in section 2.4.3, the empirical literature on real-world data for regulatory decisions does not focus on the methodological question of internal validity of effect estimates obtained from such data. Real-world data are typically associated with non-randomised studies, and methodological research has shown some open questions regarding the validity of effect estimates from non-randomised as compared to RCTs. Key findings from the

methodological literature on non-randomised studies are summarised in section 2.3.1 of the literature review.

In summary, the literature presented in section 2.4 documents an increasingly important role of non-randomised studies for regulatory approval. While regulators acknowledge the methodological limitations of non-randomised studies, the empirical literature shows that the proportion of approvals based on non-randomised studies is increasing over time, and that non-randomised studies account for the majority of pivotal trials of newly authorised medicines in some areas (i.e. cancer). Methods to address increased risk of bias do not appear to be consistently applied in non-randomised studies submitted for regulatory approval. Furthermore, there is interest in increased use of non-randomised studies that use real-world data.

2.5. Evidence standards for regulatory approval

2.5.1. Evidence standards in regular and expedited approval pathways

A body of literature exists which has analysed the evidence accepted under expedited approval pathways empirically, documenting reliance on studies measuring surrogate endpoints and non-comparative (i.e. single-arm) trial designs for approval, with continuing weaknesses in study designs and outcomes measured in post-marketing studies. This section of the literature review provides an overview of evidence standards for regulatory approval and the empirical research on how these are implemented.

Regular or "traditional" approval requires comprehensive evidence on a medicine's efficacy and safety (as well as quality, which is outside the scope of this thesis). Expedited pathways, on the other hand, provide provisions for regulators to accept more uncertainty, e.g. through acceptance of different endpoints and early results of clinical trials. However, approval under expedited pathways still requires demonstration of a positive treatment effect. Evidence standards for approval are based on both statutory requirements and longstanding regulatory practice, which are briefly explained in the section below, followed by a review of empirical studies on regulatory practice in relation to acceptance of uncertainty in expedited and regular approval pathways. Findings from these studies are described below and summarised in Box 6.

Box 6: Summary of literature review on evidence standards in regular and expedited approval pathways

What evidence standards exist for regulatory approval?

- In both the EU and US, RCTs are the preferred study design for demonstrating clinical benefit of new medicines but deviations from this standard are considered acceptable when RCTs are not feasible or not ethical.
- Some expedited pathways explicitly set different evidence standards; however, regulatory flexibility also exists in regular approval pathways.

How has regulatory acceptance of uncertainty been studied in the literature?

- Approvals under expedited pathways have been shown to be generally more likely to be based on studies measuring surrogate endpoints, uncontrolled studies, and smaller studies compared to regular approvals.
- However, both in the US and in Europe, regular approvals based on limited clinical evidence (i.e. early and/or uncontrolled trials measuring surrogate endpoints) have been documented, in particular in oncology.

Research gaps

- A comparative analysis of the EMA's and FDA's use of expedited pathways and other regulatory tools for managing uncertainty is missing.
- While expedited approval pathways are typically associated with more uncertain evidence, previous research has not focused on acceptance of uncertainty for regular approval.
- Regulatory acceptance of single-arm trials has not been analysed comprehensively for the EU.

Evidence standards in the EU

In the EU, full (or "regular") marketing authorisations are governed by Regulation (EC) No 726/2004 and Directive 2001/83/EC, stating that only products with properly or sufficiently demonstrated quality, safety, and efficacy are allowed to be made publicly accessible, and that the therapeutic benefits must outweigh potential risks (i.e. requiring a positive benefit-risk ratio). Directive 2001/83/EC specifies the contents of marketing authorisation applications which form the basis of the evaluation conducted by the EMA and are required to include results from all relevant clinical trials, both favourable and unfavourable. The Directive also includes a general preference for RCTs in Annex I, Part I, Section F:

"In general, clinical trials shall be done as 'controlled clinical trials' and if possible, randomized; any other design shall be justified. The treatment of the control groups will vary from case to case and also will depend on ethical considerations; thus it may, in some instances, be more pertinent to compare the efficacy of a new medicinal product with that of an established medicinal product of proven therapeutic value rather than with the effect of a placebo.

As far as possible, and particularly in trials where the effect of the product cannot be objectively measured, steps shall be taken to avoid bias, including methods of randomization and blinding."

While there is therefore a clear preference for RCTs, other study designs are also acceptable in specific situations. For example, evidence may come from a study with external control arm when the treatment effect is sufficiently large and the natural course of disease is well-known.[16] For cancer medicines, the EMA further specifies that "[i]n these cases, prospective confirmation in randomized, reference-controlled studies is not only unacceptable to investigators, patients and ethics committees, but also unnecessary."[177, p31]

The legal provisions do not specify what endpoints are acceptable for determining a positive benefit-risk balance. General, internationally harmonised guidance states that the "primary endpoint should be capable of providing clinically relevant and convincing evidence related to the primary objective of the study" and that the endpoints "should be meaningful for the intended population and may also take into account the views of patients".[17, p17] More detailed regulatory guidance on appropriate clinical trial design, including endpoint selection, exists for cancer medicines which states that "[t]here should thus be sufficient evidence available demonstrating that the chosen primary endpoint can provide a valid and reliable measure of clinical benefit in the patient population described by the inclusion criteria." [177, p23] The EMA guidance for evaluating cancer medicines further expresses preference for survival duration as the most convincing outcome to demonstrate patient benefit, including overall survival as well as the surrogate endpoints of progression-free survival, event-free survival, and disease-free survival. It stipulates that prolonged progression-free survival can be considered a clinically meaningful benefit because lack of disease progression is

assumed to generally translate into lack of or delay in symptom onset, worsening of quality life, and need for subsequent treatments.[177]

In the EU, special regulatory pathways accommodate more uncertainty by acknowledging that evidence may be less complete than usually required, with expectation of confirmatory evidence in the post-marketing phase (for conditional marketing authorisation) or acceptance that more robust evidence may never materialise (approval under exceptional circumstances).[178, 179] Conditional marketing authorisation requires that the benefit-risk ratio is positive based on the preliminary evidence submitted, and that the applicant is able to submit confirmatory evidence post-approval. Differently from the accelerated approval pathway in the US (see below), conditional marketing authorisation is not tied to the type of endpoint in the pivotal trial; the pathway allows acceptance of preliminary data when the benefits of immediate availability of the product are deemed to outweigh the risks of the incomplete evidence (see section 4.1.2 for more details on conditional marketing authorisation).

Evidence standards in the US

In the US, regular approval of new medicines is governed by section 505(d) of the Federal Food, Drug, and Cosmetic Act (FDCA) which stipulates that "substantial evidence" on a drug's effect is required for approval. "Substantial evidence" is defined in section 505(d) (21 U.S.C. § 355(d)) of the FDCA as follows:

> *"[E]vidence consisting of adequate and well-controlled investigations, including clinical investigations, by experts qualified by scientific training and experience to evaluate the effectiveness of the drug involved, on the basis of which it could fairly and responsibly be concluded by such experts that the drug will have the effect it purports or is represented to have under the conditions of use prescribed, recommended, or suggested in the labeling or proposed labeling thereof."*

Traditionally, this was interpreted by FDA as evidence from two independent RCTs. However, a single RCT may also be acceptable, and this flexibility was later codified in law. Moreover, FDA has issued revised guidance to industry on demonstrating substantial evidence on effectiveness which specifically addresses regulatory flexibility and situations where longstanding practices may not be

applicable.[18] While FDA guidance describes RCTs as the primary basis for substantial evidence, it explicitly includes other study designs, including studies with external controls.

Different evidence standards apply to some of the FDA's expedited pathways. While the statutory standard for effectiveness in the US does not prescribe specific endpoints, they must be considered clinically meaningful by the FDA.[18] The standard for what endpoints are considered acceptable for regular approval is set by contrasting it to the provisions for accelerated approval. Differently from regular approval, the accelerated approval pathway allows approval based on a surrogate endpoint that is considered reasonably likely to predict clinical benefit.[19] The accelerated approval pathway therefore allows interpretation of the statutory requirement for demonstrating clinical effectiveness on the basis of a non-validated surrogate endpoint, conditional on confirmation of the clinical benefit in post-marketing studies. Accelerated approval also provides FDA with the power to withdraw approval if these studies fail to confirm benefit (see section 4.1.2 for more details on accelerated approval). In addition, accelerated approvals may be granted on the basis of a single phase II study, i.e. a deviation from the traditional standard of the requirement for two RCTs. The acceptance of a single phase II study also applies in other expedited pathways in the US.[18, 76] Moreover, US legislation allows FDA to accept a single study as basis for approval (independent of any expedited pathways) if this is deemed sufficient to establish effectiveness by the agency.[98]

Design characteristics of pivotal trials in expedited pathways

As described in sections 1.4 and 1.5, regulatory bodies have implemented approval pathways that aim to accelerate the availability of new medicines by shortening the time to develop and / or review the evidence required for marketing authorisation. While these pathways may or may not formally change statutory evidence standards, they provide regulators with the scope to accept more uncertainty. Several studies have documented that approvals under expedited pathways are more likely to be based on studies measuring surrogate endpoints, uncontrolled studies, and smaller studies compared to regular approvals. Key literature on specific design characteristics of pivotal trials is presented in the sections of the literature focusing on these (namely, the use of single-arm trials with or without external controls in section 2.4.1 and 2.4.2,

observational studies in section 2.4.3, and surrogate endpoints in section 2.3.2). Below, findings of these and other studies regarding their relevance for expedited approval pathways are briefly summarised.

Overall, non-randomised studies have been documented to be more commonly used in expedited pathways compared to regular approvals in the US. Non-randomised studies more commonly formed the basis of approvals for products with orphan drug designation and for accelerated approvals in the US,[37] with an increase in the proportion of approvals through expedited pathways based on non-randomised studies over time.[36] The majority of cancer medicines with a pivotal single-arm trial were accelerated approvals, as opposed to regular approvals.[28] Approval based on non-randomised studies was also common for other expedited pathways, with more than one third of medicines with breakthrough therapy designation from 2012 (the year of inception for this pathway) to 2018 approved based on non-RCT data.[137]

In Europe, non-randomised studies have been documented as basis for both expedited and regular approvals. In cancer, the majority of approvals based on single-arm trials from 2010 to 2019 were conditional marketing authorisations,[31] but only 39% of cancer medicine approvals with externally controlled trials from 2016 to 2021 were conditional.[122]

While non-randomised studies have therefore been documented as pivotal trials for both expedited pathways and regular approvals, previous studies have not further investigated the implications of regular approval (i.e. without statutory requirements for confirmatory post-marketing evidence) based on limited evidence.

In the US, surrogate endpoints are more commonly used under expedited approval pathways compared to regular approvals.[36, 37, 142] However, approximately half of pivotal trials for approvals without any expedited pathway designation also measured surrogate endpoints and this proportion has increased over time.[36] Use of surrogate endpoints under accelerated approval has been documented to be particularly common in cancer.[61, 150, 180] Surrogate endpoints have also been shown to be most commonly used type of outcome in expedited pathways in Europe, with 90% of products with conditional marketing authorisation or accelerated assessment approved from 2011 to 2018 based on surrogate endpoints.[144]

The use of real-world evidence may be more common in expedited pathways. Among new cancer medicines approved by the FDA from 2015 to 2020 using real-world evidence to support efficacy, approximately three quarters were accelerated approvals and had breakthrough therapy designation.[132] For Europe, Eskola et al. also reported a higher proportion of products with conditional marketing authorisation using real-world evidence to demonstrate therapeutic benefit compared to products with regular approval.[176] Another study also found real-world evidence use to be commonly associated with expedited approval pathways in the US and Europe.[181] Real-world evidence features also more commonly in post-marketing obligations for medicines with conditional approval compared to regular approvals,[123] indicating a potentially important role for real-world data in regulatory practice to manage uncertainty.

Overall, the findings of the literature review on applied evidence standards show that EMA and FDA may consider the uncertainty stemming from limited clinical evidence differently; in some cases, limited evidence is sufficient for regular approval while in other cases, conditional approval is granted which requires confirmation of benefit through subsequent studies. However, comparative analyses of the acceptance of non-randomised trials and trials measuring surrogate endpoints between regulatory agencies have largely been missing from the literature, indicating an important research gap to better understand regulatory handling of uncertainty. Accordingly, chapter 4 of this thesis aims to address this gap empirically. Since that study was conducted, other analyses comparing regulatory outcomes for medicines with limited evidence have been published.[182, 183]

Acceptance of uncertain evidence for regular approval

The acceptance of more uncertain evidence is an important characteristic of expedited approval pathways that seek to accelerate time to approval for new medicines. Accordingly, the evidence base underpinning approvals through these pathways shows methodological weaknesses compared to traditional standards for regular approvals, including pivotal single-arm trials and studies measuring non-validated surrogate endpoints. However, these design characteristics are not restricted to special regulatory programmes. Both in the US and in Europe, regular approvals based on limited clinical evidence have been documented, in particular in oncology.

Overall, acceptance of uncontrolled studies was documented to be relatively common in a cohort study of all approvals by EMA and FDA without RCT data from 1999 to 2014.[157] Among 44 applications, only one was not granted approval by the FDA and nine were not granted approval by the EMA. However, the study did not analyse regulatory pathways for approvals based on non-randomised studies. Other studies have provided evidence on acceptance of non-randomised studies for regular approval. In the US, one third of cancer drug approvals with a pivotal single-arm trial were regular approvals (and the rest were accelerated approvals).[28] In the EU, the proportion of regular approvals using single-arm trials may be lower, with one study reporting expedited pathways for 16 of 22 cancer medicines authorised based on single-arm trials from 2010 to 2019,[31] and another reporting that all products for treatment of solid tumours with evidence from single-arm trials received conditional marketing authorisation.[29]

Acceptance of uncontrolled pivotal trials in oncology also means acceptance of surrogate endpoints for regular approval, since the standard outcome for measuring patient benefit – overall survival – is a time-dependent outcome that requires contextualisation through a concurrent control group. Indeed, in a cohort study of cancer medicines approved by the FDA on the basis of response rate (rather than a clinical outcome) from 2006 to 2018, more than one third received regular approval and the rest accelerated approval.[142] While accelerated approval explicitly allows for the use of non-validated surrogate markers, response rate as a marker of tumour response is also used for regular approvals.

2.5.2. Cumulative evidence on patient benefit

An important limitation of expedited pathways is the extent to which the cumulative evidence generated across pre- and post-approval settings can demonstrate clinical benefit. While some uncertainty about the clinical benefit is accepted in expedited pathways, conditional approval in particular foresees that confirmatory evidence is provided. However, empirical studies have shown that this evidence is often limited in terms of study design, selection of endpoints, and timely availability of results. Findings from these studies are described below and summarised in Box 7.

Box 7: Summary of literature review on cumulative evidence about patient benefit

How have post-marketing studies been used in regulatory practice?

- Post-marketing studies to confirm clinical benefit were shown to be commonly interventional studies, underway at time of marketing authorisation, and measuring surrogate endpoints.
- Post-marketing studies have typically been used in expedited approval pathways but may also feature in regular approvals.

What key issues with confirmation of clinical benefit have been shown in the literature?

- The methodological features of post-marketing studies result in confirmatory evidence that may not be substantially more robust than early clinical trials measuring surrogate endpoints.
- For the majority of products approved based on surrogate endpoints, no confirmatory evidence demonstrating benefit for overall survival or quality of life emerges.

Research gaps

- The role of post-marketing studies as regulatory tool to manage uncertainty has not been comprehensively evaluated.
- There is a lack of studies analysing potential differences in how regulatory bodies apply evidence standards across the product life cycle (i.e. including pre- and post-marketing phases).

Design characteristics of post-marketing studies

Research on design characteristics of post-marketing studies required by EMA and FDA has revealed that these tend to be interventional when intended to confirm efficacy (although in many cases not RCTs), commonly underway at time of marketing authorisation, and measuring surrogate endpoints, rather than patient-relevant outcomes. Finally, post-marketing studies are commonly associated with expedited pathways but are not uncommon for regular approvals.

A review of pivotal trials and confirmatory studies for drugs with FDA accelerated approval from 2009 to 2013 found that these often had similar design characteristics, including similar proportions with randomised (40% of pivotal trials vs. 56% of confirmatory studies) or uncontrolled designs (63% vs. 44%), and measuring similar endpoints (overall survival: 0% of pivotal trials vs. 6% of confirmatory studies; response

rate: 70% vs. 50%) as primary endpoint, although confirmatory studies more often measured progression-free survival (39%) compared to pivotal trials (3%).[61] The continuing reliance on surrogate endpoints was also reported in a study of cancer medicines approved by the FDA based on response rate, which found that only 21% of products with accelerated approval had later demonstrated overall survival benefit.[142] Another study of cancer medicines with FDA accelerated approval from 2011 to 2023 found that, among other characteristics, the continued use of the same trial for initial approval and confirmation of benefit and the use of primary endpoints other than overall survival or progression-free survival were statistically significantly associated with timely completion of post-marketing requirements.[184] The FDA specified overall survival as primary endpoint in only 3% of post-marketing studies.

Post-marketing studies imposed by the EMA as part of its conditional approval pathways were characterised by Hoekman et al.[74] Analysing 21 conditional marketing authorisations from 2006 to 2014, they reported that the majority of confirmatory studies were interventional (73%) and focusing on clinical efficacy (45%), and only a small minority (8%) were observational. Most of the interventional studies were ongoing at the time of approval (77%). The study did not report on endpoints used in the post-marketing studies. The EMA also published a report on its experience with conditional marketing authorisation.[156] The report describes 77 post-marketing obligations for 28 conditionally approved products. 17% of obligations had a safety primary endpoint and 12% had overall survival as primary endpoint. In more than one third of obligations, the same or similar endpoints as in the pivotal trial were used.

Post-marketing obligations are not restricted to expedited approval pathways. In a study of post-marketing studies for all cancer medicines approved by the EMA from 2004 to 2014, 80% of all approvals had at least one post-marketing obligation, although these were more common among the products approved through an expedited pathway.[185] All products with conditional marketing authorisation, approval under exceptional circumstances, and accelerated assessment had post-marketing obligations. Post-marketing obligations were more common when there were methodological limitations in the pivotal trial, including when the pivotal trial was not randomised and did not measure overall survival as primary endpoint. However, the study showed that these limitations were not necessarily addressed through post-marketing studies, as

most of them (77%) were uncontrolled and only 14% used overall survival as primary endpoint.

Post-marketing studies can also be conducted without regulatory requirement. Zeitoun et al. identified studies for medicines approved by both EMA and FDA from 2005 to 2010 that were conducted after regulatory approval (defined as trial start date within one year prior to the first regulatory approval).[186] Post-marketing studies were identified for all approved medicines in that period (median 55 studies per approved medicine). 89% of the studies were interventional, and 73% were labelled as randomised on clinicaltrials.gov, although the study authors did not report on the comparative design of the studies (i.e. whether they aimed to isolate the treatment effect compared to placebo or another intervention, or whether randomisation was between study arms of the same intervention). 63% were sponsored or co-sponsored by industry and 60% were conducted in the originally approved indication.

Evidence on confirmed clinical benefit

Given the design characteristics of post-marketing studies, the literature suggests that confirmatory evidence may not be substantially more robust than early clinical trials measuring surrogate endpoints in many cases. As a result, the available evidence on new medicines may not be sufficient to understand their benefits to patients. Indeed, evidence on patient-relevant outcomes is often missing even years after market approval, indicating that regulatory acceptance of uncertainty at time of approval is not fully resolved post-marketing.

In a study of products with FDA accelerated approval from 2009 to 2013, Naci et al. identified completed confirmatory studies for only 50% of required clinical post-marketing obligations.[61] Only one of the confirmatory studies measured overall survival while all others measured a surrogate endpoint. 83% had demonstrated some benefit for these surrogate endpoints. For a larger sample of 93 cancer drugs with accelerated approval from 1992 to 2017, Gyawali et al. reported that 62% had confirmed clinical benefit, which was based on a surrogate endpoint for 41% (20% using the same surrogate endpoint as the pivotal trial and 21% a different surrogate) and based on an improvement in overall survival for 20%.[155] The pattern of continuing reliance on the same surrogate endpoints was also documented by another study of cancer medicines with FDA accelerated approval which found that there were more conversions from

accelerated to regular approval based on a continuing use of response rate as trial endpoint (24%) than overall survival (21%).[142] The majority of conversions from accelerated to regular approval (55%) were based on demonstrated benefit on progression-free survival.

The lack of confirmatory evidence on improvement or extension of life may not preclude conversion of conditional into regular approval. Indeed, in the US, the majority (60%) of conversions from accelerated approval to regular in cancer were based on studies demonstrating benefit for surrogate endpoints and 40% were converted based on a demonstrated overall survival benefit.[154] The same study analysed overall regulatory outcomes for 46 cancer drug indications with accelerated approval, reporting that 63% had been converted to regular approval over a minimum five-year follow-up period (including one case of conversion despite a negative confirmatory trial), 15% remained under accelerated approval, and 22% had been withdrawn from the market.

For the EU, comprehensive analyses of ultimate regulatory outcomes for products with conditional marketing authorisations focus on the early years of the implementation of this pathway. In its ten-year report on the experience with conditional marketing authorisation since its introduction in 2006, the EMA reported that by the end of 2016, 14 of 28 (50%) products had been converted or were pending conversion to regular approval.[156] Many of the post-marketing obligations used for converting to regular approval used safety endpoints or continued to use the same endpoints as in the pivotal trials, and only 12% had overall survival as primary endpoint. The findings of Banzi et al., focusing on approvals from 2006 to 2015, were in line with that report, but they also highlighted that two conditionally authorised medicines had been withdrawn from the market for commercial reasons.[70] Another study analysed how often post-marketing obligations imposed by the EMA were completed, and how often they were delayed.[185] Among products with ongoing or delayed obligations, 80% were nevertheless converted to permanent marketing authorisation, indicating that regulators were satisfied with the available evidence.

Generating robust post-marketing evidence is important for understanding clinical benefit. As more evidence emerges, treatment effects observed in pivotal trials at the time of approval may be confirmed or refuted. A study of cancer drug indications

approved by the EMA from 2006 to 2016 used the ESMO-MCBS scale to assess their clinical benefit at the time of approval and up to three years post-approval.[187] Ratings were based on early and follow-up results of the same (pivotal) trials. Early results (i.e. results at the time of EMA approval) suggested a substantial benefit for 29% of 55 trials for cancer medicines with non-curative intent, intermediate benefit for 46%, and low benefit for 26%. After three years' follow-up, the proportions had increased for both substantial (to 36%) and low benefit (to 29%), showing that changes to clinical benefit ratings are likely in either direction and that regular re-assessments are necessary.

The lack of methodologically robust evidence prior to approval, and continuing into the post-marketing setting, creates challenges for clinical decision making. Without robust evidence, physicians find it difficult to appraise the therapeutic value. For example, in a study of 21 single-arm trials used as pivotal evidence for EMA approval of cancer medicines from 2012 to 2021, only three were assigned an ESMO-MCBS score indicating substantial added benefit.[29] However, challenges in determining the clinical benefit of new medicines are not restricted to approvals based on uncontrolled studies. In a cohort study of all cancer medicines approved by the EMA from 2009 to 2013 (most of which were approved with a RCT as pivotal trial), only 35% had demonstrated an overall survival benefit at the time of approval and 10% had demonstrated an improvement in quality of life.[85] The uncertainties about overall survival and quality of life were only addressed for a minority of products in the post-marketing setting, resulting in a situation where only half of the approved medicines had demonstrated prolonged or improved survival after a median period of more than five years. Extending the observation period, Grössmann et al. found a meaningful clinical benefit for one third of cancer medicines approved by the EMA from 2009 to 2020 based on the ESMO Magnitude of Clinical Benefit Scale (MCBS).[188]

Questions about the added therapeutic value of newly approved medicines extend beyond cancer. In a study of all new medicines approved by the EMA and FDA from 2007 to 2017, ratings of their added therapeutic value as determined by four national HTA bodies or authorities and one international organisation were analysed.[189] While the study found that products approved through an expedited pathway were statistically significantly more likely to be rated as having high added therapeutic value for both EMA and FDA, only 31% of all products obtained such a rating. The majority of

products with FDA expedited approval, FDA regular approval, and EMA regular approval did not have high added therapeutic value, although the study did find that two thirds of products with EMA expedited approval had high added therapeutic value. A follow-up study focusing only on products with accelerated approval in the US or conditional marketing authorisation in the EU found that the majority of products in these two pathways did not have high added therapeutic value.[190]

2.6. Identification of research gaps and research questions

2.6.1. Summary of literature review

In summary, the empirical literature on the evidence base for both expedited and regular approval highlights potentially substantial uncertainty about the clinical benefit of newly authorised medicines stemming from methodological feature of pivotal trials and post-marketing studies. Studies from both Europe and the US have demonstrated an increasing reliance on uncontrolled studies for regulatory approval, acceptance of poorly or non-validated surrogate endpoints, limited application of advanced methods for analysing and comparing data from uncontrolled studies, and an interest in using observational real-world data for establishing effectiveness. These trends in study design and data sources may contribute to uncertainty about the existence and magnitude of a treatment effect. The literature also documents that uncertainties about clinical benefit persist in the post-marketing phase, in particular for cancer medicines.

At the same time, acceptance of uncertainty allows earlier market approval and is therefore an important feature of expedited pathways. Some of this uncertainty may be addressed, including through confirmatory trials as well as an expanded use of real-world data to generate evidence on clinical benefits after initial approval. However, there are important concerns regarding both the implementation of confirmatory studies and the methodological robustness of alternative study designs to confirm clinical benefit. Against a background of empirically grounded concerns about the demonstrated clinical benefit of newly approved medicines, it is important to understand the sources and scope of uncertainty, in particular in relation to the use of non-randomised studies, and how regulators manage uncertainty.

2.6.2. Research gaps

The literature review has identified research gaps in regulatory science in relation to the methodological assumptions underpinning the trend towards accelerated approvals through early clinical evidence as well as how the uncertainty stemming from such evidence is managed. Research gaps were identified in Boxes 1 through 7 and are summarised below.

Internal validity of effect estimates from non-randomised studies

A key assumption underpinning early approval of new medicines based on non-randomised studies is that they provide valid estimates of treatment effects, or that the uncertainty about the validity of effect estimates is considered acceptable. However, existing studies comparing effect estimates from randomised and non-randomised studies have not come to a clear conclusion whether and to what extent non-randomised studies provide systematically different effect estimates than RCTs.

As shown in the literature review, regulatory agencies are aware of methodological limitations of non-randomised studies and guidance documents generally only consider them acceptable sources of evidence when more robust studies are not feasible or ethical. However, against the background of high expectations regarding the potential value of routine data (so-called real-world data) and methodological advancements in the analysis of data from non-randomised studies, there is increased interest in these study designs to support early medicines regulatory approval. As described in section 2.3, evidence from non-randomised studies may feature in the initial approval (through single-arm trials with or without external controls, which in turn may be based on real-world data, other observational data, or previous clinical trials) and in the post-marketing setting (as non-interventional studies as well as interventional non-randomised trials). Non-randomised studies for regulatory decision making may therefore feature different design, data, and analytical elements.

Empirical evidence on the internal validity of different types of non-randomised studies is needed to understand how these studies contribute to uncertainty. However, previous meta-epidemiological research was often restricted to individual therapeutic areas or methodological features. There are open questions regarding the internal validity of effect estimates obtained from the different types of non-randomised studies, data used in non-randomised studies, and analytical methods.

Regulatory management of uncertainty

The literature review has revealed a trend towards accelerating regulatory approval through the use of early clinical evidence, including non-randomised studies measuring surrogate endpoints. This trend introduces additional uncertainty regarding the clinical benefit of new medicines at the time of first market entry, and regulators need to find ways to manage this uncertainty. However, empirical studies that consider the range of regulatory tools for managing medicines with limited clinical evidence at the time of approval are lacking.

Empirically, the use of early evidence has been shown particularly clearly for cancer, and mostly for the US, with comparatively few studies focusing on Europe. The literature has mostly concentrated on characteristics of pivotal trials, highlighting design features like (lack of) randomisation and control group and the use of clinical vs. surrogate endpoints. Some studies also investigated design characteristics of post-marketing requirements but these were typically limited to approvals through expedited approval pathways, where regulators explicitly acknowledge limits in the evidence base and can use post-marketing obligations to address them. As shown in the literature review, clinical evidence with methodological weaknesses is not restricted to expedited pathways, as regular approvals in both Europe and the US also feature non-randomised studies and surrogate endpoints.

Regulatory acceptance of uncertainty may therefore exist both in expedited pathways and for regular approvals. Previous research, however, has focused on individual pathways and the use of specific regulatory tools as starting points for empirical investigations, rather than considering the options available to regulators once a limited evidence package is presented to them. These range across the product life cycle and may include refusal to grant approval if the evidence is considered insufficient, approval through an expedited pathway which allows more uncertain evidence to be accepted, requiring post-marketing studies with more or less robust designs to confirm clinical benefit (including clinical trials as well as non-randomised studies using real-world data), and potentially other regulatory tools. Different regulatory bodies may use these tools differently, revealing different approaches to regulatory acceptance of uncertainty. The breadth of regulatory tools for managing

uncertainty in the evidence base at the time of initial marketing authorisation, and how these are used by different regulatory bodies, has not been studied comprehensively.

Regulatory acceptance of non-randomised studies

Given their methodological challenges, regulatory guidance suggests a limited role for non-randomised studies, and for deviations from the traditional standard of RCTs for regulatory approval to be well-justified. Yet, previous studies have not empirically investigated how regulatory acceptance of non-randomised studies for initial marketing authorisation is justified, and whether the reasons provided for accepting such evidence are valid.

Empirical evidence presented in the literature review points towards increasing reliance on single-arm trials (with or without external controls) as pivotal evidence for regulatory approval of new medicines. The review of regulatory guidance documents suggests that these study designs are only acceptable under specific circumstances, and that RCTs are generally preferred as pivotal trials. However, in cancer, the majority of new medicines are now approved based on single-arm trials, suggesting that the exception has become the norm. Such deviation from the traditional evidence standard would be expected to be well-justified, but the literature review has not revealed studies investigating the reasons put forward by regulators or pharmaceutical companies to justify approval based on a single-arm, rather than a randomised and controlled, trial.

Moreover, the methodological limitations of single-arm trials also mean that uncertainty about the clinical benefit are likely to persist until robust evidence becomes available. While some studies have empirically investigated the timelines and characteristics of confirmatory trials following expedited approval in the US, there are no empirical studies specifically assessing whether any RCTs emerge for indications that were initially approved based on single-arm trials.

There is therefore a research gap in relation to the regulatory acceptance of single-arm trials and the justifications for relying on these study designs, rather than RCTs.

2.6.3. Thesis objectives and research questions

Thesis objectives

As described in the literature review, uncertain or immature evidence on new medicines is often associated with expedited approval pathways, which can change evidence requirements or the time available for regulatory review, although uncertainty can also feature in regular approvals. An important determinant of uncertainty is clinical study design. Specifically, non-randomised studies are at increased risk of bias compared to RCTs and can therefore introduce substantial uncertainty about the clinical benefits of new medicines when used as pivotal evidence to support regulatory approval. The use of surrogate endpoints, rather than patient-relevant outcomes, can help shorten evidence generation, but introduces additional uncertainty about the translation of surrogate measures into clinical benefit. Uncertainty in relation to study design and endpoints is compounded when single-arm trials (with or without external controls) provide pivotal evidence for regulatory approval. At the same time, non-randomised evidence may be used to address open questions at the time of initial marketing authorisation by generating evidence on benefits and harms in the post-marketing setting. Both when used as pivotal evidence and when used to generate additional evidence in the post-marketing setting, methodological concerns regarding the validity of effect estimates obtained from non-randomised studies exist.

This thesis focuses on the uncertainty inherent to non-randomised studies as evidence base for new medicines and investigates how this is handled by regulatory bodies. It addresses the policy question of how regulatory bodies balance rigorous evidence standards with the goal of facilitating patient access to new medicines and contributes to the methodological question of the internal validity of effect estimates obtained from non-randomised studies.

Specifically, this thesis aims to investigate the methodological assumptions underpinning the increased interest in non-randomised studies for regulatory approval of medicines and how regulators manage uncertainty about clinical benefit stemming from methodologically less robust pivotal trials. This overarching aim is addressed through three specific research questions.

Research question 1

Traditionally, regulatory approval of new medicines has been based on evidence generated in RCTs which represent the commonly accepted gold standard for measuring benefits and harms of health interventions. Deviating from this standard may introduce additional uncertainty into regulatory decision making. Compared to well-conducted RCTs, non-randomised studies are subject to risk of bias and may produce systematically different effect estimates. Against the background of increased interest in the use non-randomised studies to either produce evidence on a new medicine's benefits and harms before market approval, or to confirm the treatment effect in the post-marketing setting, the first research question investigates the internal validity of effect estimates obtained from non-randomised studies:

How similar are treatment effects for new medicines obtained from non-randomised studies compared to RCTs?

Research question 2

As regulatory bodies seek to balance their mandate to safeguard public health with early market access for promising new treatments, they may accept more uncertainty in the clinical evidence. This may be introduced through evidence that would regularly not be considered sufficient for regular approval, such as non-randomised phase II trials measuring surrogate endpoints. The relatively immature nature of the evidence may be addressed through different regulatory tools, including the use of conditional approval pathways that require confirmatory evidence to be produced once the product is on the market, mandating the generation of post-marketing evidence more generally, restricting the authorised indication to patients where more certain evidence exists, and refusal to grant marketing authorisation.

The second research question addresses the different tools that regulatory bodies can use to manage uncertainty, including uncertainty introduced through the use of non-randomised studies:

Which tools are used by regulatory bodies to manage uncertainty in the clinical evidence when assessing new medicines for market approval, and how do they impact on the cumulative evidence available on benefits and harms of new medicines?

Research question 3

While the methodological rigour of well-conducted RCTs is undisputed, there may be valid reasons for deviating from this standard and to instead rely on non-randomised studies for regulatory approval. However, RCTs have been shown to be feasible even in challenging conditions, such as for rare diseases, raising the question of how the trade-off between relatively more certain evidence (through RCTs) and relatively less certain, but earlier evidence (through non-randomised studies) is justified. The third research question therefore investigates the reasons for regulatory acceptance of non-randomised studies and their validity:

What are the reasons for regulators to accept non-randomised studies for marketing authorisation of new medicines, and are these reasons justified?

3. Meta-epidemiological study of treatment effects in randomised and non-randomised studies of medicines

Part of the work presented in this chapter was published in *JAMA Network Open*.[191] Compared to the published manuscript, this chapter was edited, and it provides additional details on background, methods, and results, and an expanded discussion section.

Chapter summary

While RCTs have long been considered the gold standard for evaluating the clinical efficacy and safety of health interventions, there is increasing interest in using non-randomised studies for regulatory approval of medicines. It is therefore important to understand whether non-randomised studies produce different effect estimates for pharmacological interventions compared to RCTs. This chapter presents findings from a meta-epidemiological study of 346 meta-analyses that included both randomised and non-randomised studies addressing the same clinical question.

Across all 346 meta-analyses with 2,746 contributing individual studies, there was no strong evidence for a systematic difference in treatment effects obtained from non-randomised vs. randomised studies. However, the effect estimates obtained from the two study designs indicated different statistical conclusions about whether a medicine provided benefits or harms for 38% of meta-analyses, and disagreements about the size of the treatment effect were beyond chance for 16%. There was evidence for systematically more favourable effect estimates from experimental (i.e. interventional) non-randomised studies compared to RCTs.

Although there was no overall difference in effect estimates between study types, non-randomised studies both over-estimated and under-estimated treatment effects observed in randomised studies and introduced additional uncertainty. The findings presented in this chapter therefore indicate that replacing RCTs with non-randomised studies may lead to substantial uncertainty about the effects of new medicines.

3.1. Introduction

3.1.1. Role of RCTs and non-randomised studies for medicines regulation

In the last 60 years, RCTs have widely been regarded as the gold standard for assessing the clinical efficacy and safety of new medicines.[192, 193] As described in section 1.2.1, RCTs minimise bias by randomly assigning treatments to participants and therefore provide regulatory bodies, payers, clinicians, and patients with valid evidence on what treatments work. In contrast to RCTs, treatment assignment in non-randomised studies is influenced by the patient, the provider, or the setting. Non-randomised studies are therefore particularly susceptible to bias due to confounding, which arises from systematic distortions in the true treatment effect caused by uneven distribution of participant characteristics between intervention and control groups, and to selection bias, which arises when the characteristics of individuals who enter or remain in a study are related to both treatment assignment and outcomes (see also sections 2.3 and 2.4 for a discussion on methodological challenges in relation to non-randomised studies).[15] Consequently, discrepancies may emerge between the results of RCTs and non-randomised studies.

The question of whether RCTs and non-randomised studies yield comparable results has far-reaching implications. In the past, evidence from non-randomised studies suggesting a beneficial effect has led to patients being exposed to drugs that were later proven to be ineffective in RCTs. Examples include hydroxychloroquine for treating severe COVID-19, off-pump coronary-artery bypass graft, and beta-carotene for reducing cancer risk.[194-196]

Importantly, and as described in more detail in section 1.2.2, the internal validity of non-randomised studies has recently attracted renewed interest due to a growing enthusiasm for incorporating non-randomised studies into decision-making processes throughout the lifecycles of new medicines. This enthusiasm is fuelled by increasing availability of electronic health record data, administrative data, and innovative methods for analysing these datasets. In response, medicines regulatory agencies in the US and Europe are actively exploring the feasibility and validity of utilising non-randomised studies in their decision-making processes. The 21st Century Cures Act in the US has granted the FDA the authority to consider non-randomised studies when

assessing the efficacy and safety of medicines for supplemental indications.[73] In Europe, the EMA is undertaking initiatives to create a regulatory environment for using data from non-randomised studies,[35] and the proposal for a revision of the EU's pharmaceutical regulation suggests that data collected outside of clinical trials (so-called "real-world data") are regarded as a key aspect of regulatory decision making going forward.[197] While non-randomised studies have been traditionally used as a complement to RCTs, there is now a growing interest in potentially substituting or replacing RCTs with well-conducted non-randomised studies.[32]

3.1.2. Previous research comparing randomised and non-randomised studies

Previous research has examined the comparability of treatment effects between RCTs and non-randomised studies, yielding varied findings, as summarised in the literature review of this thesis (section 2.3.1).[95, 96, 109-111, 198-201] However, the most recent comprehensive review, encompassing 45 clinical questions and 408 individual studies, was published more than 20 years ago.[95] Most published studies focused on selected therapeutic areas, limiting the generalisability of their findings. Meta-analyses investigating treatments for COVID-19, incorporating both observational studies and RCTs, generally reported consistent summary treatment effects.[202] Most recently, replication studies for highly selected clinical questions with good data availability have identified a general alignment between RCTs and their non-randomised emulations, although disagreements in results were observed in approximately one fourth of the cases.[48] A comprehensive review of potential discrepancies between treatment effects of randomised and non-randomised studies among a non-selected sample of clinical questions from the 21st century is missing.

3.1.3. Study aims

The primary objective of this study was to assess and compare treatment effects of the same medicine when evaluated in non-randomised studies (NRS) versus RCTs.

3.2. Methods

This was a comprehensive meta-epidemiological analysis encompassing 2,746 studies across 346 distinct clinical topics. The general analytical framework for meta-epidemiological research comparing effect estimates for studies with different characteristics is described in Box 8. The specific methods used in this study, including the range of measures for discrepancy in treatment effects obtained from NRS and RCT employed, are described below. A protocol for this study was registered on the PROSPERO database of systematic reviews (CRD42018062204).

Box 8: Modelling discrepancies between results obtained from different study types

In meta-epidemiological research, the results of two sets of studies – one with a characteristic of interest and the other without it – are compared to understand the potential impact of study characteristics on effect estimates.[107, 203] For example, a set of high-quality studies *(H)* could be compared with a set of low-quality studies *(L)*. Quality could be operationalised through an overall risk of bias score or through a specific study characteristic indicative of high quality, such as the use of random allocation of participants to intervention and control groups. All other characteristics being equal, the average difference between treatment effects obtained from *H* versus *L* is therefore an estimate of the systematic discrepancy in effect estimates (i.e. bias) introduced through the characteristic that distinguishes *H* from *L*. In the meta-epidemiological study presented in chapter3, potential discrepancies in effect estimates obtained from randomised vs. non-randomised studies are analysed.

For any given study, *i*, an estimate of the treatment effect can be written as

$$\theta_i = \delta_i + \beta_i X_i$$

Where δ_i is the component of the estimate representing the true treatment effect, β_i is an estimate of potential bias introduced through a given study characteristic, and X_i is an indicator that takes on value 1 for studies where that study characteristic is present and 0 otherwise. In the example of high- and low-quality studies above, X_i=1 for studies from group *L* and X_i=0 for studies from group *H*.

Bias in a given study, β_i, is assumed to follow a normal distribution, with mean *b* and variance κ^2.

If the bias is associated with a study characteristic, this is expected to be present regardless of the specific clinical question studied. Comprehensive collections of all relevant studies answering a clinical question can be thought of as meta-analyses. Within a given meta-analysis, *m,* potential bias in a given study, *i,* can be written as

$$\beta_{im} \sim Normal(\beta_m \kappa^2)$$

The parameter κ^2 represents variation of bias within a given meta-analysis, *m*. A high value for κ^2 indicates uncertainty with respect to where in the distribution of potential bias the estimate for a given study *i* within that meta-analysis would fall.

If the bias is present across different collections of studies, or meta-analyses, it can be written as

$$b_m \sim Normal(b_0, \varphi^2)$$

Here, φ^2 represents variation of bias across meta-analyses. A high value therefore indicates uncertainty with respect to where in the distribution bias would fall for a given meta-analysis (or collection of studies).

3.2.1. Identification of clinical questions

The study was conducted for a set of clinical questions for which meta-analyses including at least one RCT and one NRS were conducted to obtain estimates of the effectiveness of pharmacological treatments, as defined in the participants, interventions, comparators, outcomes (PICO) framework. Clinical questions with potentially eligible meta-analyses were identified through three sources: (1) a database search in MEDLINE (via PubMed) for existing meta-epidemiological studies comparing RCTs and NRS, (2) a database search in MEDLINE (via PubMed) for systematic reviews including both RCTs and NRS, (3) a review of all systematic reviews indexed in the Cochrane Database of Systematic Reviews that included both RCTs and NRS. Only records published from 2009 to 2018 were included to cover clinical questions from the last decade before the meta-epidemiological study was conducted. Originally, the study protocol foresaw covering meta-analyses published from 2000 to 2018 but this was restricted to 2009 to 2018 due to feasibility constraints. Details of the database searches are available in Appendix 2.

Meta-analyses were only eligible for inclusion when RCTs and NRS contributed to a single pooled estimate of the effectiveness of a pharmacological treatment, following the *"within meta-analyses"* approach for meta-epidemiological studies.[204] This approach capitalises on the subject matter expertise of researchers and clinicians conducting meta-analysis in their area of interest, and who judged RCTs and NRS to be sufficiently similar to each with other with respect to study participants, intervention, comparator, and outcome, to provide evidence on an intervention's benefits or harms..

Systematic reviews where RCTs and NRS were meta-analysed separately (e.g. due to concerns about heterogeneous patient populations in the two study designs) were excluded.

Potential source systematic reviews containing such meta-analyses, as identified through database searches, were screened at the title and abstract level independently by two reviewers. Conflicting decisions were resolved by consensus. Full texts of remaining records were screened by one reviewer, after double screening of a 10% sample of records showed almost perfect agreement ($\kappa = 0.85$).

For each included source systematic review, one meta-analysis was selected for data extraction. Data extraction was conducted for the meta-analysis of the primary outcome. In cases where the meta-analysis of the primary outcome did not include both RCTs and NRS, the next most prominently presented outcome with the highest number of contributing RCTs and NRS was selected. All identified meta-analyses and individual studies were compared on the basis of unique identifiers (preferably PubMed IDs, and in case not available ad hoc created IDs) to identify possible overlap.[205] Within each meta-analysis, only unique individual studies were eligible to contribute effect estimates. The same individual study was eligible to contribute to several meta-analyses.

3.2.2. Data extraction

Meta-analysis-level and study-level information was extracted from source systematic reviews using a pre-specified spreadsheet by a single researcher. A guidebook with instructions for each item was prepared and data extraction was checked by a second researcher for approximately 10% of meta-analyses. Where possible, pre-specified categories were used for extracting study design characteristics, including for types of NRS and RCTs and data sources, among others. Appendix 3 provides details on the data extracted from source systematic reviews.

Categorisation of study designs was based on typologies used in previous meta-epidemiological reviews.[109, 199] RCTs and NRS were distinguished, with the former defined by the use of a random sequence to allocate study participants to intervention and control groups, and the latter by the absence of such random sequence. Categorisation of studies into RCTs and NRS was based on the assessment made by the

authors of the source meta-analyses, except for studies included in Cochrane systematic reviews, where the detailed and consistently applied and reported risk of bias assessment was used to distinguish studies using truly random vs. quasi-random sequences (e.g. date of birth, social insurance number) for participant allocation.

For NRS, further distinctions were made between experimental (or interventional) and observational (or non-interventional) designs, a categorisation also applied by scholars, medicines regulators, and other stakeholders.[12, 109, 171, 206] This distinction was also implicitly referred to in a checklist for quasi-experimental studies.[207] Interventional or experimental NRS are studies in which the investigator has some control over study conditions, including the allocation of participants into treatment and control groups. Examples include quasi-randomised trials (where the allocation mechanism falls short of true randomisation, see above) and non-randomised clinical trials (where allocation is by patient or physician preference). Single-arm trials with external controls as described in the literature review of this thesis (sections 2.4.1 and 2.4.2) are examples of experimental NRS. By contrast, observational or non-interventional NRS lack the experimental intention of experimental NRS. They exploit natural variation in the use of interventions to study patient outcomes. Observational cohort studies as described in the literature review of this thesis (section 2.4.3) are examples of non-interventional NRS. Although the terms interventional and non-interventional study are more commonly used in the regulatory context, experimental and observational NRS, respectively, are used in this study to reflect the broader interest in the internal validity of study designs which goes beyond regulatory use cases.

Other study-level information was extracted as reported in the source meta-analyses, including analytical methods used, data sources of NRS, participant details, intervention details, risk of bias tool and score, number of participants in intervention and control arms, effect measure, and a measure of its variance (see Appendix 3).

3.2.3. Analysis

Conversion of effect estimates

All effect estimates were converted into log odds ratios (OR) and coded so that an OR<1 indicated a beneficial effect of the novel intervention.

For studies reporting risk ratios (RR), hazard ratios (HR), or incidence ratios (IR), it was assumed that these would approximate the OR if the baseline risk was below 0.2.[208] This assumption was checked by extracting a baseline or assumed control risk for each of the meta-analyses reporting RR, HR, or IR. The assumed control risk was selected as reported by the authors of the source systematic reviews (e.g. more recent Cochrane systematic reviews include a summary of findings table with assumed control risk).[209] Where the source review did not report a baseline risk, the event rate of the control group of the largest NRS for that topic was selected, which was assumed to represent the best approximation of a clinically relevant population. For meta-analyses reporting continuous outcomes, these were first converted into standardised mean differences (SMD)[210] and then to OR,[211] following the approach taken in previous meta-epidemiological research covering multiple therapeutic areas.[212] Conversion of continuous outcomes into OR was required for 72 meta-analyses (20.8% of all included meta-analyses).

The practice of coding (or "coining") of effect estimates in meta-epidemiological research has been the subject of intense methodological debate following the publication of a study comparing effect estimates from non-randomised studies with those of subsequently published RCTs.[96, 213] This debate focused on potential bias in the calculation of summary estimates of discrepancies between ORs obtained from two study types when the coding of the direction of effect (e.g. a beneficial effect being represented by OR<1) is systematically based on the observed results of one of the two study types (in the case of the meta-epidemiological study by Hemkens et al., non-randomised studies published before the first RCT became available [96]). Since the sequence of evidence generation was not the focus of the meta-epidemiological study presented in this chapter, coding of effect estimates was not systematically based on one of the two study types but on a pre-specified contrast of novel vs. old intervention. For meta-analyses comparing two active interventions, the novel intervention was identified through the descriptions provided by the authors of the source systematic review or through web searches in cases where the novel intervention could not be identified with certainty from the source systematic review. In a sensitivity analysis, only meta-analyses with placebo comparisons were included where the expected direction of effect was unambiguous.

Measures of discrepancy between estimates from NRS and RCTs

Descriptive analyses included first plotting the summary estimates for NRS and RCTs conducted for the same clinical question and reporting the number of meta-analyses for which the NRS and RCT effects, respectively, were more favourable. Within each meta-analysis, summary estimates and 95% confidence intervals (CIs) of NRS and RCTs, respectively, were calculated using a random effects Hartung-Knapp-Sidik-Jonkman meta-analysis model to take into account between-study heterogeneity.[214, 215] Potential discrepancies between treatment effects obtained from NRS and RCTs were analysed using four measures.

First, the frequency of substantial disagreement in summary ORs was measured by counting how often the summary OR obtained from one type of study was at least twice as favourable as the other (i.e. the OR obtained from one study type was at least half the OR obtained from the other study type).[95]

Second, the frequency of discrepancies in the summary logOR beyond chance at the 5% significance level was measured.[95] Summary logORs for NRS and RCT for each meta-analysis were compared using

$$log_{ROR} = \log(OR_{NRS}) - \log\,(OR_{RCT}),$$

and a 95% CI was computed using standard error (SE) of logROR,

$$SE(log_{ROR}) = \sqrt{SE(OR_{NRS})^2 + SE(OR_{RCT})^2},$$

to then compare these CIs with the null value of logROR = 0.

Third, the number of meta-analyses for which the summary estimates of NRS and RCTs, respectively, led to different statistical conclusions was reported. Different statistical conclusions included cases where the meta-analytic estimate for NRS and RCTs, respectively, showed contradictory and statistically significant treatment effects (i.e. meta-analyses for which the summary OR for NRS was statistically significantly <1 while the summary OR for RCTs was statistically significantly >1, and vice versa), and where one study type produced a statistically significant meta-analytic effect estimate (i.e. summary OR entirely <1 or entirely >1) while the estimate for the other study type remained inconclusive (i.e. 95% CI of the summary OR included 1).

Fourth, discrepancies between NRS and RCTs at the aggregate level were quantified through a two-stage meta-analysis to obtain ratios of OR (ROR) for treatment effects obtained from NRS vs. RCTs.[107] The analysis was implemented in a Bayesian framework, with uninformative priors for the discrepancy of treatment effects between NRS and RCTs.[216] The variation of discrepant treatment effects between NRS and RCT results across meta-analyses was also quantified using the between-meta-analysis standard deviation in discrepancies (ϕ) and the variation of discrepancies across studies within meta-analyses using the between-study standard deviation in discrepancies (κ).[203, 217] These measures indicate variation in effect estimates obtained from different study designs: higher values indicate a wider spread in the magnitude of discrepancies between two study types across meta-analyses (ϕ) and across individual studies within meta-analyses (κ).

Other measures for assessing discrepancies in treatment effects exist, such as correlation and concordance coefficients and the absolute ROR.[95, 96, 198, 200, 202, 203, 218-220] This study focused on measures that were deemed important from a clinical or regulatory decision-making perspective, i.e. that provide estimates of both absolute and relative discrepancies, potential differences in statistical conclusions drawn, and direction of deviation.

Analyses were implemented in Stata version 13.1 (StataCorp) and WinBUGS version 1.4.3 (Imperial College and Medical Research Council).

3.2.4. Subgroup and sensitivity analyses

Subgroup analyses were conducted for pre-specified characteristics at the meta-analysis level (including type of comparator, median risk of bias of included studies, and type of outcome) and study level (including type of NRS, analytical method used in NRS). Additional subgroup analysis to explore heterogeneity in the discrepancy in treatment effects in RCTs vs. NRS was conducted by data source of NRS, type of control in NRS, therapeutic area, how narrowly a meta-analysis was defined in the PICO framework (and therefore, how well aligned RCTs and NRS included in the same meta-analysis were), and methodological quality of source meta-analyses. Study-level characteristics were often not reported in detail by source meta-analyses, resulting in

small sample sizes for most subgroups. The study therefore focuses primarily on results of subgroup analyses for the following characteristics:

- Type of NRS: The discrepancy in treatment effects between experimental NRS and RCTs may be different from observational NRS, which are not typically conducted in a controlled clinical research setting.
- Type of comparator: In studies with an active comparator, the novel intervention may not be clearly identified, and it may therefore be unclear in which direction bias would operate.[213] On the other hand, a clear direction of bias operating in favour of the intervention can be expected in meta-analyses comparing a pharmacological intervention to placebo or no treatment.
- Type of outcome: Mortality represents the most objective outcome and is commonly deemed the most important outcome for health interventions. Studies measuring subjective outcomes may be at higher risk for bias overall, and the impact of the lack of random allocation of participants may be obscured by differential reporting of subjective outcomes in RCTs vs. (observational) NRS, in which blinding would not be feasible. By stratifying the analysis according to outcome, a more homogenous sample was created to focus on the association of randomisation with treatment effects.
- Matching of RCTs and NRS in a meta-analysis by PICO criteria: Scores from 1 (low) to 3 (high) were assigned to each meta-analysis based on how narrowly participants, intervention, comparator, and outcome were defined in each topic to give an overall "narrowness index" (range 4-12).
- Methodological quality: The sample was restricted to meta-analyses that were considered to be likely the most methodologically robust, using two criteria: meta-analyses published in the journals with the 10 top 5-year impact factors as of January 2020, and reviews conducted by Cochrane groups.

A priori specified subgroup analyses included the type of analytical method employed and for the use of real-world data (RWD) in NRS. These results are based on a limited number of topics which are not likely to be representative.

- Type of analysis used in NRS: Where this information was available from source meta-analyses, NRS using naïve analysis (no adjustment), some form of adjustment for patient characteristics (one or more covariates included in

regression models), propensity score-based analysis, and matching by individual patient characteristics was analysed to assess heterogeneity in the agreement of effect estimates according to the analytical methods used.

- Data source used in NRS: Where this information was available from source meta-analyses, NRS were grouped according to whether data were collected specifically for the study, studies that used clinical RWD (disease registries and case records), and studies that used non-clinical, administrative RWD (administrative or claims data).

In a post-hoc sensitivity analysis, the sample was restricted to meta-analyses where NRS were published before the first RCT to account for potential influence of known RCT results on the design, conduct, and reporting of NRS.[96]

3.3. Results

A total of 10,957 records were screened at the title and abstract level, and 830 were reviewed in full, resulting in a total of 336 included records (Figure 2). These 336 records contributed 346 unique meta-analyses (two meta-epidemiological studies contributed more than one meta-analysis), with 2,746 contributing individual studies (median 3 RCTs [range 1-92] with median 379 participants [interquartile range 147-1,162] per meta-analysis [median 100 participants per RCT, range 5-235,600, interquartile range 50-245] and median 2 NRS [range 1-44] with median 416 participants [interquartile range 100-2,301] per meta-analysis [median 195 participants per NRS, range 6-2,145,593, interquartile range 73-1,041]).

3.3.1. Sample characteristics

Summary characteristics of included meta-analyses are presented in Table 2. Details of included meta-analyses are presented in Appendix 4. Included meta-analyses ranged across a wide range of therapeutic areas, with anti-infectives system use (n=66, 19.1% of included meta-analyses) and blood and blood forming organs (n=64, 18.5%) the most common categories. Roughly two thirds of topics had a placebo comparator (n=226). The quality of the match between RCTs and NRS was assessed as high for 111 (32.1%) meta-analyses, moderate for 166 (48.0%), and low for 69 (19.9%). The median publication year for individual studies was 2010 or later for 159 (46.0%) meta-

analyses. For 146 (42.2%) meta-analyses, the first RCT was published after NRS were already published.

Figure 2: Flow chart of selection of meta-analyses for meta-epidemiological study

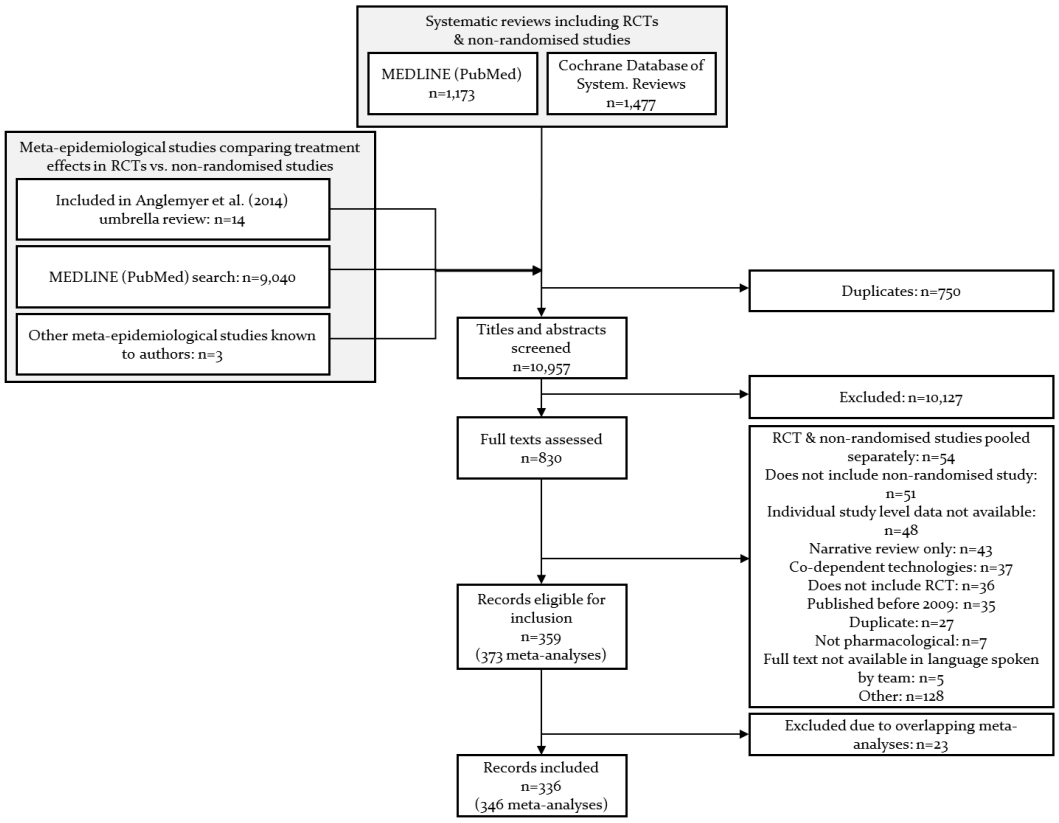


Table 2: Characteristics of meta-analyses including both non-randomised and randomised studies

Characteristic	Number of meta-analyses (%)
All meta-analyses	346 (100)
Comparator	
Active	94 (27.2)
Placebo/no treatment	226 (65.3)
Includes both active and placebo-controlled studies	26 (7.5)
Outcome type	
Mortality	59 (17.1)
Other objective outcome	158 (45.7)
Subjective outcome	126 (36.4)
Includes studies with different types of outcomes	3 (0.9)
Therapeutic area by WHO ATC first level categorisation	
Antiinfectives for systemic use	66 (19.1)
Blood and blood forming organs	64 (18.5)
Cardiovascular system	45 (13.0)
Antineoplastic and immuno modulating agents	43 (12.4)
Nervous system	27 (7.8)
Alimentary tract and metabolism	23 (6.6)
Systemic hormonal preparations	19 (5.5)
Genito-urinary system and sex hormones	14 (4.0)
Other categories combined	45 (13.0)
Risk of bias across NRS in a meta-analysis	
Low median risk of bias	97 (28.0)
Moderate median risk of bias	61 (17.6)
High median risk of bias	124 (35.8)
No risk of bias information	64 (18.5)
Risk of bias across RCTs in a meta-analysis	
Low median risk of bias	91 (26.3)
Moderate median risk of bias	95 (27.5)
High median risk of bias	103 (29.8)
No risk of bias information	57 (16.5)
Time period of studies included in the meta-analyses	
Median publication year pre-2000	56 (16.2)
Median publication year 2000-2009	131 (37.9)
Median publication year 2010 and later	159 (46.0)
Matching quality of RCTs and NRS in meta-analysis	
High (score of 10-12 out of 12)	111 (32.1)

Moderate (score of 7-9 out of 12)	166 (48.0)
Low (score of 4-6 out of 12)	69 (19.9)
Timing of evidence generation	
NRS published before first RCT	146 (42.2)
First RCT published before NRS	169 (48.8)
First NRS and first RCT published in the same year	31 (9.0)

Legend:

Abbreviations: NRS, non-randomised study(ies); RCT, randomised controlled trial(s); WHO ATC, World Health Organization Anatomical Therapeutic Chemical classification system.

Risk of bias across NRS (RCTs) in a meta-analysis: the table shows the proportion of meta-analyses for which the median of the risk of bias scores of NRS (RCTs) included in that meta-analysis was low, moderate, or high.

Matching quality of RCTs and NRS in meta-analysis: the table shows the proportion of meta-analyses for which the quality of the matching between NRS and RCTs included in the meta-analysis was deemed high, moderate or low according to how closely aligned each of the four PICO components (participants, intervention, comparator, outcome) were between NRS and RCT. A score from 1-3 was assigned for each of the four PICO components according to how well NRS and RCTs included in the same meta-analysis were matched.

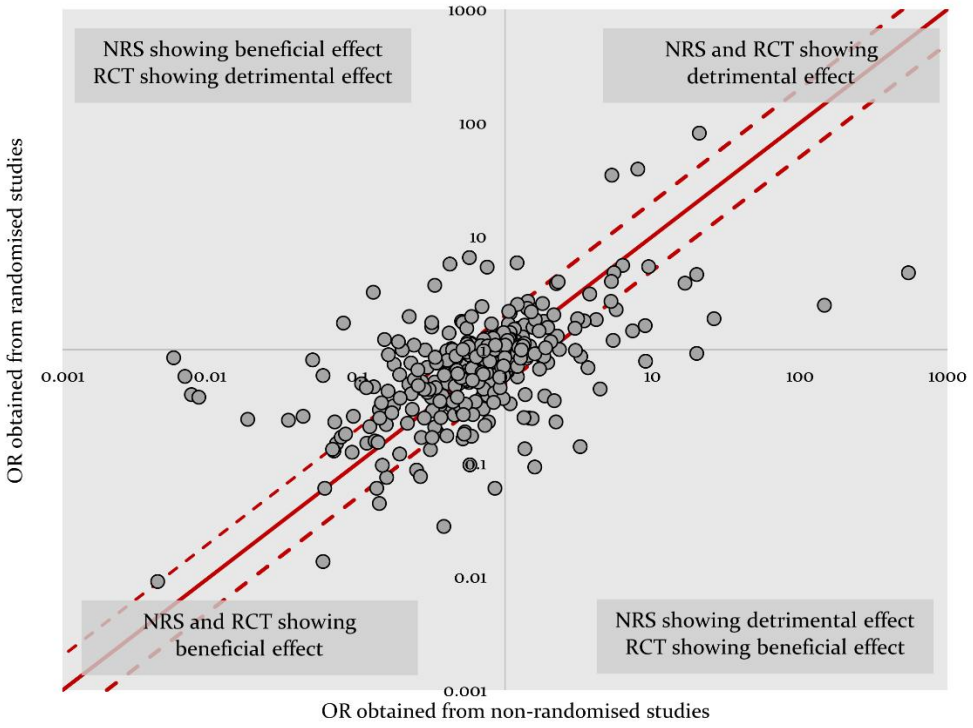
3.3.2. Descriptive results for variation in effect estimates

Variation in discrepancies between treatment effects is displayed in Figure 3, which shows summary effect estimates obtained from RCTs and NRS for all 346 meta-analyses. NRS gave a more favourable effect (i.e. a lower summary OR) for 186 (53.8%) meta-analyses, and RCTs gave a more favourable effect for 158 (45.7%) meta-analyses. For 121 (35.0%) meta-analyses, the summary OR obtained from one study type was twice as large or more (i.e. half the OR, or less) as from the other (Table 3), including 65 (18.8% of all meta-analyses) where NRS indicated a substantially more beneficial effect and 56 (16.2%) where RCTs indicated a substantially more beneficial effect. Disagreement between study types was beyond chance for 54 (15.6% of all) meta-analyses, including 30 (8.7%) where the summary OR obtained from NRS was more beneficial, and 24 (6.9%) where the summary OR obtained from RCTs was more beneficial (Table 3).

In a subgroup analysis that only included experimental NRS, these gave a more favourable effect than RCTs for 74 meta-analyses (60.7% of all meta-analyses including experimental NRS). For 55 meta-analyses (45.1% of all meta-analyses including experimental NRS), the summary OR from one study type was twice as favourable as the other (Table 3), including 36 (29.5%) where the OR obtained from experimental NRS was half the OR of RCTs or less. Disagreement between study types was beyond

chance for 31 (25.4%) meta-analyses with experimental NRS. Subgroup analysis of meta-analyses with observational NRS showed more favourable effects in NRS than RCTs for 120 meta-analyses (52.9% of all meta-analyses including observational NRS). The summary OR were twice or half as large in one study type vs. the other for 40 (17.6%) meta-analyses, and disagreement was beyond chance for 31 (13.7%).

Figure 3: Agreement of summary effect estimates (odds ratios, OR) obtained from randomised and non-randomised studies for 346 meta-analyses



Legend:

Each circle shows the summary OR obtained from a meta-analysis of randomised (vertical axis) and non-randomised (horizonal axis) studies for one clinical question. OR<1 indicates a beneficial effect. The solid red line indicates perfect agreement (exact same summary OR obtained from randomised and non-randomised studies) and dashed red lines indicate substantial disagreement (OR obtained from randomised studies is at most half of the OR obtained from non-randomised ones, or vice versa). Circles in the upper left quadrant show meta-analyses where NRS evidence indicates a beneficial effect (summary OR<1) and RCT evidence a detrimental effect (summary OR>1), and circles in the bottom right quadrant show meta-analyses where NRS evidence indicates a detrimental effect (summary OR>1) and RCT evidence a beneficial effect (summary OR<1). Circles in the upper right quadrant show meta-analyses where both NRS and RCT evidence indicate a detrimental effect: circles above the red line indicate a larger detrimental effect in RCTs and circles below the red line indicate a larger detrimental effect in NRS. Circles in the bottom left quadrant show meta-analyses where both NRS and RCT evidence indicate a beneficial effect: circles above the red line indicate a larger beneficial effect in NRS and circles below the red line indicate a larger beneficial effect in RCTs.

Abbreviations: NRS, non-randomised study(ies); OR, odds ratio; RCT, randomised controlled trial(s).

Table 3: Results for four measures of discrepancy between non-randomised studies and RCTs

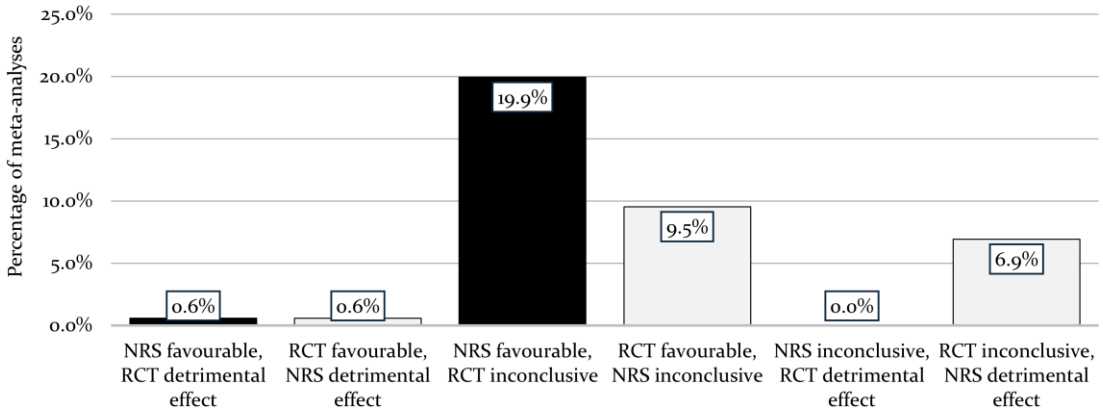
	Measure of discrepancy between non-randomised studies and RCTs			
	Summary OR twice as favourable for one study type vs. the other (frequency (%))	Discrepancy in summary OR beyond chance (frequency (%))	Meta-analyses with different statistical conclusions (frequency (%))	ROR (95% CrI)
Full sample (n=346 MA)	121 (35.0)	54 (15.6)	130 (37.6)	0.95 (0.89–1.02)
NRS study type				
Experimental NRS (n=122 MA)	55 (45.1%)	31 (25.4%)	42 (34.4%)	0.81 (0.67–0.97)
Observational NRS (n=227 MA)	40 (17.6%)	31 (13.7%)	89 (39.2%)	0.98 (0.87–1.06)
Outcome type				
Mortality outcome (n=59 MA)	13 (22.0%)	11 (18.6%)	28 (47.5%)	0.95 (0.84–1.08)
Other objective outcome (n=161 MA)	50 (31.1%)	32 (19.9%)	56 (34.8%)	0.95 (0.85–1.06)
Subjective outcome (n=126 MA)	50 (39.7%)	37 (29.4%)	46 (36.5%)	0.97 (0.84–1.09)
Comparator				
Active comparator (n=108 MA)	40 (37.0%)	19 (17.6%)	32 (29.6%)	0.91 (0.76–1.07)
Placebo / no treatment comparator (n=237 MA)	43 (18.1%)	38 (16.0%)	97 (40.9%)	0.95 (0.88–1.04)
Matching quality of RCTs and NRS in meta-analysis				
Good match RCT-NRS (n=111 MA)	35 (31.5%)	24 (21.6%)	40 (36.0%)	0.91 (0.79–1.03)
Moderate match RCT-NRS (n=166 MA)	53 (31.9%)	37 (22.3%)	57 (34.3%)	0.98 (0.87–1.12)
Poor match RCT-NRS (n=69 MA)	19 (27.5%)	19 (27.5%)	33 (47.8%)	0.96 (0.87–1.08)
Methodological quality of source meta-analyses				
Top journals only (n=118 MA)	48 (40.7%)	27 (22.9%)	45 (38.1%)	0.88 (0.76–1.03)
Cochrane reviews only (n=78 MA)	36 (46.2%)	16 (20.5%)	29 (37.2%)	0.83 (0.65–1.03)

Timing of evidence generation				
NRS published before first RCT (n=146 MA)	53 (36.3%)	31 (21.2%)	50 (34.2%)	0.95 (0.83–1.08)

Abbreviations: CrI, credible interval; MA, meta-analyses; NRS, non-randomised study(ies); OR, odds ratio; RCT, randomised controlled trial(s); ROR, ratio of odds ratios

RCTs and NRS led to different statistical conclusions about the therapeutic benefit of pharmacological interventions in 130 (37.6%) meta-analyses, based on comparing 95% CIs around the OR from either study type with a null effect (Figure 4). The most common category of discrepant statistical conclusions was NRS showing a favourable effect while evidence obtained from RCTs was inconclusive (n = 69, 19.9% of meta-analyses). In 33 (9.5%) meta-analyses, RCTs showed a favourable effect while the NRS evidence was inconclusive. Contradictory treatment effects were observed in four (1.2%) meta-analyses, including two where the effect estimate obtained from NRS indicated a statistically significant favourable effect while the RCT estimate indicated a detrimental effect and two where RCT evidence suggested a favourable effect and NRS evidence a detrimental effect.

Figure 4: Discrepancies in statistical conclusions about therapeutic benefit of pharmacological interventions based on evidence obtained from non-randomised studies or RCTs



Legend:

Figure shows proportions of meta-analyses based on the conclusions about the existence of a therapeutic benefit drawn from NRS or RCT evidence. A favourable or detrimental effect was deemed to exist if the 95% confidence interval of the summary OR did not include 1. Evidence was considered inconclusive if the 95% confidence interval of the summary OR included 1.

Abbreviations: NRS, non-randomised study(ies); RCT, randomised controlled trial(s).

3.3.3. Results from the main analysis

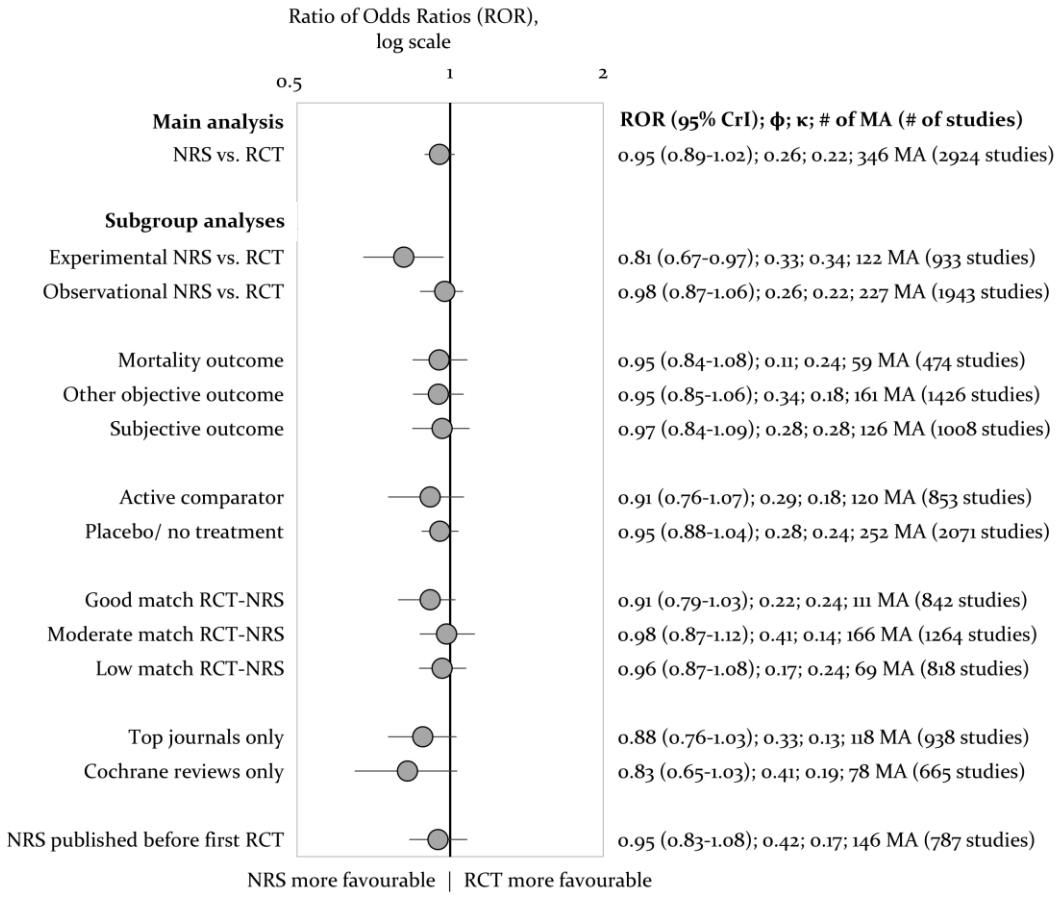
In the main analysis, there was no evidence of a difference between effect estimates obtained from NRS vs. RCTs on average when pooling discrepancies across all 346 meta-analyses (ROR 0.95, 95% CrI 0.89 to 1.02; Figure 5). In subgroup analyses, effect estimates obtained from experimental NRS were more favourable compared with RCTs

(ROR 0.81, 95% CrI 0.68 to 0.97), while no difference was observed between observational NRS and RCTs (ROR 0.98, 95% CrI 0.87 to 1.06). Stratifying meta-analyses by how narrowly defined they were (and therefore how closely aligned NRS and RCT included in the same meta-analysis were) suggests a potential trend of NRS showing more favourable results in meta-analyses with high quality matches between NRS and RCTs, and when limiting the sample to meta-analyses published in top journals or as part of Cochrane systematic reviews. However, there was no strong evidence of systematically different results for RCTs and NRS in these subgroups.

Variation in the discrepancy of treatment effects was present between studies within meta-analyses ($\kappa = 0.22$) and between meta-analyses ($\phi = 0.26$). Variation between meta-analyses was reduced for meta-analyses measuring mortality ($\phi = 0.11$) compared with other objective or subjective outcomes ($\phi = 0.34$ and 0.28, respectively). There were no systematic differences in between-meta-analysis variation (ϕ) or within-meta-analysis variation (κ) for the other characteristics at meta-analysis level.

In 146 (42.2%) meta-analyses, the first NRS was published before the first RCT. In this subset of meta-analyses, findings were consistent with the overall sample (Table 3): in 53 meta-analyses (36.3% of the 146 included in the sensitivity analysis), the summary OR was twice as favourable for one study type vs. the other; in 31 (21.2%), the discrepancy in summary OR was beyond chance, 50 (34.2%) reached different statistical conclusions; and the ROR was 0.95 (95% CrI 0.83 to 1.08; Figure 5).

Figure 5: Results of meta-meta analytic comparison



Legend:

Figure shows ratio of odds ratios (ROR) comparing effect estimates obtained from non-randomised studies to effect estimates obtained from randomised studies, and heterogeneity parameters (ϕ, between-meta-analysis heterogeneity; κ, increase in within-meta-analysis heterogeneity). Results are shown for all meta-analyses (top), followed by subgroup analyses by type of NRS; different types of outcomes; types of comparators; matching quality of RCTs and NRS in the same meta-analysis; high-quality publications; and timing of NRS and RCT.

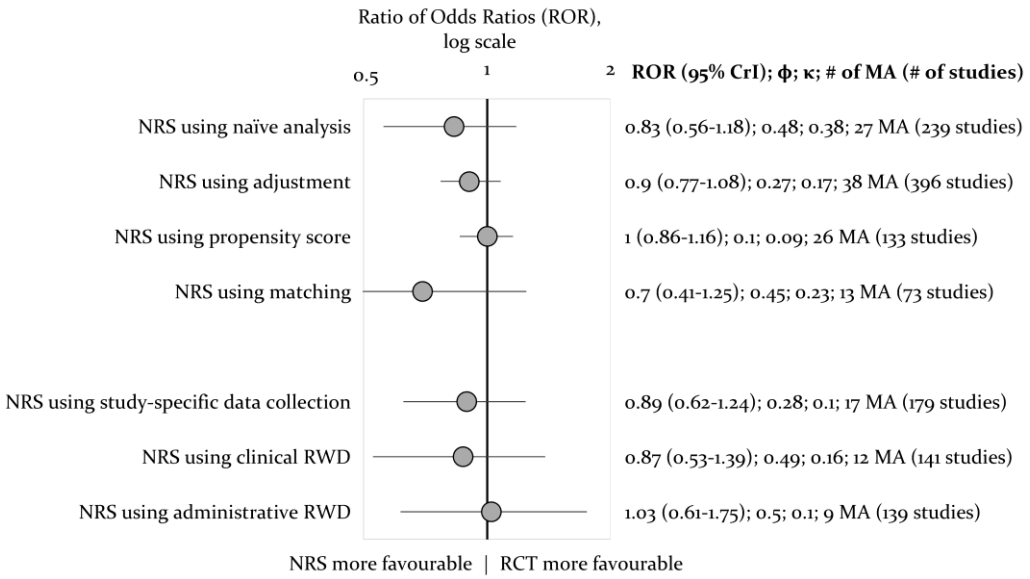
Abbreviations: MA, meta-analyses; NRS, non-randomised studies; RCT, randomised controlled trials; ROR, ratio of odds ratios.

Subgroup analysis by study-level characteristics

Study-level data regarding analytical methods and data sources used in NRS were only available for a small subset of meta-analyses (in most cases less than 10% of all included original studies, and representing less than 10% of all included meta-analyses). Interpretation of the ROR for these groups is therefore limited. Focusing on uncertainty in the discrepancy between effect estimates from RCTs and NRS showed that between-meta-analysis variation (ϕ) and within-meta-analysis variation (κ) were reduced for studies using propensity score methods compared with other analytical methods

(Figure 6). No similar pattern of reduced variation was observed for data sources in NRS.

Figure 6: Meta-meta analytic results from additional subgroup analyses for study-level characteristics



Abbreviations: MA, meta-analyses; NRS, non-randomised studies; RCT, randomised controlled trials; ROR, ratio of odds ratios; RWD, real-world data.

3.4. Discussion

3.4.1. Summary of main findings

This meta-epidemiological study examined the differences in treatment effects between randomised and non- randomised studies of pharmacological interventions across 346 clinical topics. This study represents the largest investigation of its kind to date. The analysis did not uncover any systematic under- or overestimation of treatment effects in non-randomised studies when compared to RCTs. However, this overall finding masks substantial variability in the observed differences between treatment effects derived from RCTs and non-randomised studies. A considerable number of meta-analyses exhibited discrepancies in effect estimates, including 38% where statistical conclusions about the existence of an effect differed, 35% where effect estimates differed by a factor of two or more, and 16% where disagreements about the size of the treatment effect were beyond chance. There was evidence for systematically more favourable effect estimates from experimental (i.e. interventional) non-

randomised studies compared to RCTs. This study shows that decision makers face uncertainty around both the direction and magnitude of potential disagreement between randomised and non-randomised studies: non-randomised studies both over-estimated and under-estimated treatment effects observed in randomised studies.

3.4.2. "Target trial" emulation vs. other non-randomised studies

This study extends previous research investigating the comparability of treatment effects derived from RCTs and non-randomised studies.[112] In particular, it provides generalisable findings across a broad range of therapeutic areas, and quantifies uncertainty associated with treatment effects derived from non-randomised studies. These findings are important to inform the debate about the role of non-randomised studies as potential substitutes for RCTs for regulatory approval of medicines (see section 1.2.2). Previous meta-epidemiological reviews yielded mixed results,[95, 109-111] with varying factors such as outcome types,[199] study timing,[96] and analytical methods in non-randomised studies contributing to discrepancies across reviews.[112, 200, 221, 222]

Drawing from a comprehensive sample of meta-analyses including both RCTs and non-randomised studies, there was a 38% probability of reaching different statistical conclusions regarding a medicine's effectiveness depending on the type of study design considered, and a 62% probability of reaching the same statistical conclusion. This finding broadly aligns with a recent study that sought to emulate highly selected RCTs using administrative data.[48] In the study by Wang et al., treatment effects derived from non-randomised studies yielded concordant conclusions with 56% of the RCTs they aimed to emulate. The present meta-epidemiological study used a different approach to Wang et al. and other observational studies aiming to emulate RCTs by design using the "target trial" approach.[49] This approach was described in more detail in sections 2.3.1 and 2.4.3. Briefly, in the RCT emulation literature, observational studies aim to either strictly replicate inclusion criteria, intervention details, and outcomes of existing trials or to create a hypothetical target trial in the absence of RCTs. While these studies apply advanced methodological standards to improve replicability of RCT findings with observational data, there are question marks about their routine use as basis for decision making due to limited availability of the data required to implement

them.[128] While their use is increasing,[126] target trial emulations only represent a small subset of non-randomised study designs. Other non-randomised studies are not designed to replicate RCT results, but they still matter for decision makers: when faced with a situation of uncertainty about the benefits of a treatment, decision makers, including regulators, HTA bodies, payers, and clinicians, will likely use evidence from any study design available, including different types of non-randomised studies. It is therefore important to understand potential discrepancies in effect estimates obtained from RCTs and from the body of non-randomised studies conducted.

By focusing on clinical questions where subject matter experts included both RCTs and non-randomised studies in the same meta-analysis, this study aimed to strike a balance between internal validity (striving for high comparability between RCTs and non-randomised studies conducted for the same clinical question) and external validity (looking for answers that go beyond individual clinical domains and case studies of the feasibility of implementing advanced methods).

3.4.3. Variation in discrepancies for different study designs

An important contribution to the literature lies in advancing the understanding of how discrepancies in treatment effects vary according to different types of study designs, analytical methods, and data used in non-randomised studies. While some of the previous meta-epidemiological studies focused on specific analytical methods and types of data used in non-randomised studies, this study provides novel evidence on how these perform when compared against RCTs.

In the context of regulatory approval of medicines, single-arm trials, which may or may not include explicit comparison to an external control group, play an important role, as described in sections 2.4.1 and 2.4.2. It is therefore particularly important to understand how effect estimates from non-randomised clinical trials compare to RCTs. This study found that effect estimates obtained from experimental (i.e. interventional) non-randomised studies were systematically more favourable than those obtained from RCTs (overestimating RCT estimates by 19%). Previous meta-epidemiological studies typically categorised NRS according to the use of different observational study designs, such as cohort or case-control studies,[112] but they did not specifically investigate potential differences in effect estimates for experimental non-randomised studies.

These study designs share important validity traits with RCTs, such as a controlled environment for administering the treatment and strict participant inclusion criteria, and feature increasingly prominently in regulatory approval of medicines, as described in the introduction and literature review sections of this thesis. Nevertheless, the absence of random participant allocation in these studies can introduce bias through confounding. Indeed, this study indicates that such bias may operate, leading to on average 19% more favourable effect estimates compared to RCTs. Experimental non-randomised studies showed at least twice as favourable treatment effects as RCTs for 45% of meta-analyses.

The same bias through confounding would also be expected to operate in other types of non-randomised studies. However, for observational studies, the mechanism of bias through confounding (predicting a more favourable effect for the new intervention) may be directly counteracted by "real-world" effects, which are expected to attenuate the treatment effect observed in highly controlled settings, including differences in population characteristics, the administration of the novel intervention, and different standards of usual care compared to the controlled RCT environment. In this study of 346 meta-analyses with more than 2,700 individual studies, these characteristics were not assessed at the study-level. When grouping meta-analyses according to how narrowly they defined inclusion criteria for individual studies in terms of participants, intervention, comparator, and outcomes, the most narrowly defined meta-analyses with the highest quality matching of RCTs and non-randomised studies suggested potentially larger discrepancies in effect estimates than meta-analyses with lower quality matching of RCTs and non-randomised studies. However, there was no clear evidence on differential discrepancy in treatment effects according to the quality of matching between RCTs and non-randomised studies across the four measures of discrepancy.

Similar to earlier studies,[95, 96, 109, 111, 198-200] there were no clearly identified characteristics that predict agreement between findings of RCTs and non-randomised studies. The between-meta-analysis variation in discrepancies (ϕ) remained consistent across various subgroup analyses. However, discrepancies in effect estimates were observed to be less variable when the outcome of interest was mortality, suggesting that potential bias in non-randomised studies may be less variable in such cases. Conversely,

higher variation in potential bias was found for subjective outcomes, including those assessed by clinicians.

3.4.4. Policy implications of the internal validity of non-randomised studies

Regulatory approval based on non-randomised studies

This study has important policy implications. As described in detail in the literature review section of this thesis, non-randomised studies are playing an increasingly important role in influencing decisions about the approval and reimbursement of new medicines.[121, 175, 176, 223] The use of non-randomised studies has been on the rise for some time. Between 2015 and 2017, approximately 18% of new drugs gained approval in the US based on non-randomised studies, up from just 6% between 1995 and 1997.[36] The majority of these approvals were based on single-arm trials, which lacked a control group, addressing unmet clinical needs within narrow therapeutic indications.[28, 157] Small patient numbers in such rare diseases render conducting RCTs more challenging. Importantly, medicines approved via single-arm trials may not have their efficacy confirmed in post-marketing RCTs: in a cohort study of oncology medicines approved in the US based on single-arm studies via the accelerated approval pathway between 1992 and 2020, the FDA requested confirmatory RCTs for approximately 70% of the approved indications, but RCT evidence became available for only 41%.[30] Other studies have also raised concerns over the confirmation of clinical benefit in post-marketing studies, as discussed in more detail in section 2.5.2. This may partly be due to marketing authorisation complicating patient recruitment for post-approval RCTs.[28] Consequently, drugs approved based on single-arm trials are more likely to be evaluated in non-randomised studies during the post-marketing period. While single-arm trials without external control group were not included in this meta-epidemiological study (included studies required a control arm to provide a comparative effect estimate for the source meta-analyses), its findings of potentially considerable disagreement about therapeutic benefit shown in other types of non-randomised studies and RCTs suggest that caution is warranted when relying on non-randomised studies as substitutes for RCTs.

Going forward, non-randomised studies may also play a more prominent role for initial approval, as introduced in sections 1.2.2 and 2.4.4.[26] In 2023, the FDA released

draft guidance naming observational data (RWD) as potentially suitable confirmatory evidence for regulatory approval, replacing the previously used standard of two independent clinical studies.[98] It is therefore important to understand the benefits and risks of relying on non-randomised studies for the evaluation of new medicines. Potentially considerable disagreement about therapeutic benefit between the two study types suggest that caution is warranted when relying on non-randomised studies as substitutes for RCTs.

Implications for downstream decision makers

This study has important implications for decisions about pricing and reimbursement of new medicines, as well as for clinical practice. Reimbursement decisions, and therefore routine availability of a health technology in a health system, are commonly informed by assessments of the comparative effectiveness of a new product vs. existing treatments. Methodologically, these assessments are more challenging to conduct in the absence of robust head-to-head evidence from RCTs. The methodological challenges of comparative benefit assessments for HTA and reimbursement decisions in relation to non-randomised studies are discussed in more detail in section 6.2.3.

This study also has important implications for clinical practice. Clinicians and patients have learned to rely on evidence from RCTs for informing treatment decisions. Of course, RCTs may also be at high risk of bias due to problems with their design, conduct, analysis and reporting.[14] Despite these concerns, the findings of this meta-epidemiological study underline the importance of RCTs, as the conclusions about a treatment's effect may differ when based on non-randomised studies. The statistical conclusions about a medicine's treatment effect were different for almost four in 10 clinical questions. Running non-randomised studies to identify treatment effects therefore risks exposing patients to treatments that may ultimately be proven ineffective. Indeed, medical reversals have occurred because RCTs provided conclusive evidence about the benefits and harms of long-standing medical practices that were based on evidence obtained from non-randomised studies.[194-196, 224, 225] If conclusions about the treatment effects derived from non-randomised studies are later found to be invalid, patients may lose confidence in therapeutic decisions and treatments may be interrupted, leading to adverse health outcomes. Empirical evidence

about the comparability of treatment effects obtained from RCTs and non-randomised studies is therefore an important building block for evidence-based medicine.

Improving the feasibility of RCTs to reduce uncertainty

Section 2.3 introduced the key concept of uncertainty over clinical benefit in the regulatory approval of medicines and its association with study design characteristics. This study revealed substantial uncertainty about effect estimates obtained from non-randomised studies. Similar to findings in previous meta-epidemiological research, non-randomised studies sometimes overestimated and sometimes underestimated treatment effects obtained from RCTs. Importantly, these discrepancies were large (effect estimates differing by a magnitude of 2 or more) and clinically meaningful (change in statistical conclusions about the therapeutic benefit) for more than one third of clinical questions. There was also substantial variation in discrepancies between effect estimates from the two study designs. Predictions of the extent to which effect estimates from non-randomised studies deviate from those obtained from RCTs in a given situation would therefore be subject to considerable uncertainty. These results suggest substantial uncertainty for decisions made solely on the basis of non-randomised studies, highlighting the continued need for robust RCTs.

However, the need for and feasibility of RCTs is increasingly challenged, as described in section 1.2.1. The trajectory towards increasing reliance on non-randomised studies may partly be explained by concerns about the cost and complexity of RCTs.[226] Decision makers are beginning to question whether the added complexity of RCTs justifies any validity gains they offer. Paradoxically, there appears to be a limited effort to simplify the design and conduct of RCTs in the post-marketing setting. As the push towards non-randomised studies gains more traction, it could potentially impede the necessary progress required to improve the feasibility of RCTs.[227] The findings of this study suggest that responses to concerns about cost and complexity of RCTs need to be addressed.

3.4.5. Limitations

Like all meta-epidemiological studies, this is an observational study which limits causal interpretation of results.[204] However, meta-epidemiological research has proved to be an important tool to generate empirical evidence that advances our

understanding of risk of bias in health care research. For example, the most widely used tool for assessing risk of bias in studies of health interventions, the Cochrane Collaboration's risk of bias tool for RCTs, was developed partly on the basis of empirical evidence generated by meta-epidemiological studies.[228, 229]

This study included 346 distinct clinical questions that were the subject of meta-analyses published from 2009 to 2018. Only meta-analyses where subject matter experts combined both RCTs and non-randomised studies in the same meta-analysis were included, indicating that they considered the two study types to provide relevant evidence for the clinical question they were interested in. Rather than a representative sample of all clinical questions, these meta-analyses may be more representative of those with overall limited levels of evidence (otherwise, only RCTs would be expected to be included in a meta-analysis). The clinical questions included in this meta-epidemiological study therefore provide insights into situations with substantial uncertainty about whether a medicine works. Including both study types in the same meta-analysis may also reflect limited methodological understanding of the authors of source meta-analyses but the conclusions of this study did not change when restricting the sample to meta-analyses conducted by Cochrane groups or those published in high-impact journals.

Authors of source systematic reviews may have been more likely to include both RCTs and non-randomised studies in the same meta-analysis when there was a prior expectation or knowledge about comparable evidence. The sample identification strategy applied in this meta-epidemiological study meant that clinical questions were excluded when the authors of source systematic reviews determined that there were substantial differences between the two study types – possibly due to observed differences in results. This may have resulted in an underestimation of the true difference between treatment effects obtained from RCTs and non-randomised studies for all clinical questions for which both study types exist. Analysing all clinical questions for which both RCTs and non-randomised studies exist might therefore show more pronounced differences in treatment effects.

The study used the OR as common summary effect measure across all included meta-analyses. Clinical interpretation may be limited through the conversion of continuous outcomes into OR. A common threshold for a clinically meaningful effect

could not be specified across all 346 clinical questions, resulting in a potential limitation of the clinical interpretation of observed discrepancies in effect estimates between RCTs and non-randomised studies. The study reported discrepancies according to different thresholds representing large differences irrespective of the individual clinical question.

3.5. Conclusions

In this meta-epidemiological study, discrepancies in effect estimates obtained from non-randomised and randomised studies were observed with respect to both magnitude of effect and statistical conclusions across a substantial subset of clinical questions about the therapeutic effect of pharmacological interventions. While there was overall no systematic difference in effect estimates obtained from non-randomised vs. randomised studies, experimental non-randomised studies were found to produce on average 19% larger treatment effects compared with RCTs. These findings suggest that caution is warranted when relying on non-randomised studies as substitutes for RCTs.

4. Comparative analysis of approval of cancer medicines with uncertain therapeutic value in Europe and the United States

Part of the work presented in this chapter was published in an article in *The Milbank Quarterly*.[230] In addition to editorial changes, this chapter includes additional background information, details on the methods and results, and four illustrative examples of different regulatory outcomes in Europe and the US.

Chapter summary

Aiming to speed up availability of new medicines, conditional approval pathways have been introduced in the US and Europe that allow regulatory approval on the basis of early evidence, including non-randomised studies measuring surrogate endpoints. These regulatory pathways, which require confirmatory evidence after approval, are commonly used to approve new cancer medicines. This chapter compares how EMA and FDA used regulatory tools, including conditional approval, refusal of market approval, post-marketing obligations, and restrictions in approved indications, to manage uncertainty in the clinical evidence base for 21 cancer medicine indications with a regulatory outcome by both agencies from 2009 to 2013.

The majority of pivotal trials for the 21 indications with uncertainties in the clinical evidence were single-arm trials. Although EMA and FDA mostly assessed the same trials, regulatory outcomes frequently differed, with the majority of indications receiving full (or regular) approval by one agency and conditional approval by the other. There were also important differences in the use of post-marketing studies to confirm clinical benefit: while the FDA more often required RCTs, the EMA more often requested non-randomised studies. Non-randomised evidence featured in post-marketing obligations for three quarters of EMA approvals and in one third of FDA approvals. Both agencies continued to rely on surrogate measures in the post-marketing setting and commonly faced delays in post-marketing studies.

The study presented in this chapter shows that both EMA and FDA commonly accept uncertainty when approving cancer medicines. While some differences exist in how uncertainty is managed between the two agencies, robust evidence on patient-relevant endpoints is unlikely to materialise after conditional approval in both the EU and US.

4.1. Introduction

4.1.1. Evidence standards and early market access

In the EU and the US, medicinal products regulators (EMA and FDA, respectively) are responsible for assessing the benefit-risk profile of new medicines before they enter the market. Typically, this requires developers of new medicines to submit substantial evidence from controlled clinical trials. As discussed in sections 1.6.1 and 2.3, robust study designs, in particular randomised trials, have traditionally been preferred as evidence to establish clinical benefit, and deviating from robust standards introduces additional uncertainty. Generating comprehensive data on benefits and risks takes time as new therapeutic agents pass through different phases of research and development, as described in section 1.5. In their quest to safeguard and improve public health, regulatory bodies have to balance the need for comprehensive data to make informed decisions with the potential benefits of new medicines, in particular when these address an unmet medical need. This "evidence vs. access conundrum" [63] can also be framed in terms of the willingness of regulators to accept uncertainty. On the one hand, placing a greater weight on the availability of complete evidence for regulatory decision making means more certainty that medicines available on the market have a positive benefit-risk profile. This would typically mean that evidence is obtained from RCTs measuring hard, patient-relevant endpoints such as survival. Conversely, granting approval on the basis of less complete evidence, such as a non-randomised study measuring a surrogate endpoint, shows greater willingness to accept uncertainty about a new medicine's therapeutic value.

Figure 1 in section 1.5 shows how approval on the basis of a phase II trial (rather than a confirmatory phase III trial) can contribute to earlier market access for a new product. Compared to phase III trials, phase II trials are less likely to be RCTs. By definition, they intend to provide early (rather than confirmatory) evidence on a product's efficacy and safety, and accordingly are more likely to measure surrogate endpoints. As described in sections 1.5 and 2.3.2, study design is directly linked to the selection of study endpoints: while confirmatory trials are expected to report on patient-relevant outcomes, such as overall survival, these require a concurrent control group; single-arm trials, on the other hand, can measure surrogate endpoints which

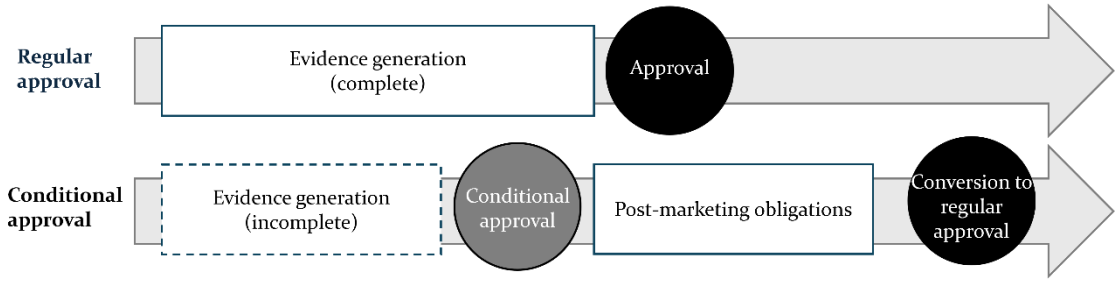
accumulate more quickly and may result in shorter trial duration, but cannot reliably generate evidence on outcomes for which a concurrent counterfactual is needed.[23]

In both the US and Europe, expedited pathways were introduced to create a regulatory framework allowing pharmaceutical manufacturers to submit preliminary (or immature) evidence about their product's efficacy and safety, deviating from the evidence standards used for regular approval.[65, 66, 76] These are most commonly used for cancer medicines both in the EU and the US.[55, 156, 231] Details on the different types of expedited pathways – targeting the time it takes to generate evidence or to subsequently review it, or allowing consultation between medicines developers and regulators prior to the authorisation process – are described in section 1.4. This study focuses on the most prominent expedited approval pathways – conditional marketing authorisation (CMA) in the EU and accelerated approval (AA) in the US – because they aim to reconcile the limitations of early evidence, such as non-randomised phase II studies, with an overall robust evidence standard over a product's life cycle.

4.1.2. Conditional approval pathways and other regulatory tools to manage uncertainty

Although some differences exist between CMA and AA (see Table 4 for a comparison of key characteristics of the two pathways), they share two key features: in both pathways, approval is initially granted on the basis of preliminary evidence not usually considered sufficient to warrant approval and approval is conditional on the submission of confirmatory evidence in the post-marketing setting. They can therefore be regarded as regulatory tools to navigate situations with substantial uncertainty about the efficacy and safety of new treatments. CMA and AA are particularly instructive for studying how regulatory bodies deal with uncertainty because they shift part of the evidence generation for regulatory decision making from the pre-approval to the post-approval period (Figure 7). Throughout chapter 4, the term "conditional approval" is used to refer to both CMA and AA, and the term "post-marketing obligations" is used to refer to post-marketing studies imposed by the regulator.

Figure 7: Schematic illustration of evidence generation in regular vs. conditional approval pathways



In principle, the conditional approval pathway illustrated in Figure 7 allows regulators to uphold regular evidence standards when considering the totality of evidence generated in the pre-approval and post-approval periods while granting early marketing authorisation to promising new treatments: approval is granted based on early evidence, but this should be substantiated through rigorous confirmatory studies. However, a body of literature is emerging which questions whether evidence standards in the CMA and AA pathways are rigorous enough to ensure patient benefit, as described in section 2.5. Research on special approval pathways has pointed towards questionable novelty of products that benefited from these programmes,[55, 232] a higher number of safety events in the US (although not in Europe),[233, 234] and an erosion of the evidence landscape, with robust evidence on the efficacy and safety of new medicines often unavailable at the time of marketing authorisation and unlikely to become available in a timely manner in the post-approval phase.[61, 62, 70, 235]

Accelerated approval

The AA pathway was introduced in the US in 1992 in response to the HIV/AIDS crisis and aimed to address important public health needs by accelerating the development and marketing of products for serious or life-threatening conditions.[3] While the AA pathway initially benefited treatments for infectious diseases (approximately 65% of AA medicines in the first decade of the programme), it is now primarily used in oncology (accounting for 83% of AA from 2012 to 2021).[236]

The pathway allows products to enter the market based on early evidence that would not be considered sufficient for regular approval (Table 4). Differently from other measures aiming to expedite development and market entry of new products

(such as priority review and breakthrough designation), it changes the evidence standard as new medicines are allowed to be approved based on surrogate endpoints that are considered reasonably likely to predict clinical benefit (i.e. surrogate endpoints that are not validated; see section 2.5.1 for a description of evidence standards under AA).[55] Clinical benefit then has to be confirmed through obligatory post-marketing studies.[237] In theory, approval may be revoked if confirmatory studies fail to demonstrate clinical benefit. However, empirical studies have shown that uncertainties in the clinical evidence at the time of AA are often not addressed through confirmatory studies, as they continue to rely on surrogate endpoints,[82, 155] and may not be completed in time.[61]

Conditional marketing authorisation

The CMA pathway was introduced in the EU in 2006. Similar to the FDA's AA pathway, it aims to accelerate the availability of treatments for seriously debilitating or life-threatening diseases by allowing confirmatory evidence on clinical benefit to be demonstrated through obligatory post-marketing studies (Table 4). The CMA pathway was initially often used retroactively, suggesting a role as "rescue option" for products that otherwise would not have been authorised.[238] However, this approach may have changed over time, as an increasing number of CMAs are requested proactively.[239]

As described in section 2.5.1, the EMA's CMA does not formally change evidence standards, as products granted CMA still need to demonstrate a positive benefit-risk-ratio. However, this can be based on incomplete data when a set of criteria are fulfilled, including that the manufacturer is able to provide confirmatory evidence in the post-marketing setting.[240] Other important differences between CMA and AA include the restricted scope of CMA to initial marketing authorisations, whereas the FDA can grant AA also for subsequent indications, and the annual review of CMAs until all post-marketing obligations are considered fulfilled by the EMA.

Similar to the FDA's AA pathway, CMA is used primarily in oncology.[156]

Table 4: Comparison of conditional approval pathways in Europe (Conditional Marketing Authorisation, CMA) and the US (Accelerated Approval, AA)

	EMA Conditional Marketing Authorisation [240]	**FDA Accelerated Approval** [237]
Key features	• Requires four criteria to be fulfilled (see below for evidence assessment), including a positive benefit-risk profile and the ability to provide comprehensive data through post-marketing studies • Marketing authorisation is reviewed every year until conditions of initial authorisation are fulfilled • Can only be granted for first marketing authorisations	• Allows market entry of medicines on the basis of a surrogate endpoint that is "reasonably likely" to predict clinical benefit • Clinical benefit is to be confirmed through post-marketing trials • Can be granted for first marketing authorisation as well as further indications for approved medicines
Eligibility criteria	Medicinal products that fulfil one of the below: • treat, prevent, or diagnose seriously debilitating or life-threatening diseases, • are used in response to public health emergency situations, • are designated as orphan medicinal products	Drugs and biologics that fulfil all of the below: • address an unmet medical need in the treatment of a serious condition, • provide a meaningful advantage over available therapies (where a therapy exists), • have been shown to be effective on the basis of an endpoint "reasonably likely" to predict clinical benefit
Evidence assessment	Requires that four criteria are fulfilled: • Positive benefit-risk ratio is established • The manufacturer is likely to be able to provide additional data	The same statutory requirements as for regular approval must be met: • Substantial evidence on efficacy based on adequate and well-controlled trials

	• The product fulfils an unmet medical need • Benefits to public health of immediate availability of the product outweighs the risks of incomplete evidence	• Sufficient information to determine the medicine is safe However, the evidence standard is different: • Efficacy can be demonstrated on the basis of a surrogate endpoint that was determined to be "reasonably likely" to predict clinical benefit • FDA accepts that evidence will generally come from fewer, smaller, and shorter trials
Theoretical consequences for failure to comply with conditions	Failure to comply with post-marketing obligations will lead to an assessment report by the EMA, which can result in one of the following:[241] • Letter to the marketing authorisation holder (i.e. the pharmaceutical company) • Oral explanation by the marketing authorisation holder • Initiation of procedure to vary, suspend, or revoke approval • Inspection	• Failure to conduct post-marketing studies "with due diligence" may result in withdrawal of approval • Violation of post-marketing obligations may also result in civil monetary penalties of up to US$ 250,000 [242]

Other regulatory tools to manage uncertainty

Both EMA and FDA have introduced conditional approval pathways to accelerate availability of new treatments for serious conditions with unmet medical needs, which are predominantly used for cancer medicines. When regular approval is not considered feasible due to limited data, conditional approval (i.e. initial approval being contingent on submission of confirmatory evidence) can be considered as one of several tools at the disposal of regulatory agencies to manage uncertainty. Other possible ways of handling uncertainty include the following:

- Denying approval when the evidence is not considered sufficient to indicate a positive benefit-risk ratio, and when the limitations in the evidence are not considered likely to be addressed through post-marketing obligations.
- Restricting the final approved indication when the evidence indicates a positive benefit-risk ratio for a subgroup of patients.
- Requiring the manufacturer to conduct post-marketing studies to address specific data limitations. This forms part of the model for conditional approvals but post-marketing obligations can also be imposed for regular approvals.

4.1.3. Existing literature and research gaps

With novel cancer medicines increasingly targeting narrow, previously underserved patient populations,[243] and clinical evidence for these typically coming from early-phase studies with less robust study designs (single-arm trials rather than RCTs, and/or measuring surrogate endpoints), situations where only limited data are available are likely to occur more regularly in the future. It is therefore important to compare the regulatory approaches adopted by the FDA and EMA to managing this uncertainty, and investigate reasons for any discrepancies. Through conditional approval and other tools, regulatory bodies have some leeway in how they manage uncertainty.[244] How the EMA and FDA use these tools to manage uncertainty has not been studied comprehensively in the literature.

Previous studies analysing regulatory practices in the approval of cancer medicines often focused on a single regulatory agency,[82, 85, 155, 245, 246] including some that reviewed how uncertainty was handled by the EMA.[238, 244] In a comparative study of EMA and FDA decisions, Kashoki et al. assessed concordance between the two agencies

across all therapeutic indications and regulatory pathways, reporting overall high concordance and identifying differences in clinical evidence submitted as important reasons for observed discrepant regulatory outcomes.[182] While the study included a subset of cancer indications and reported approval pathways, the analysis included all products with matched indications across EMA and FDA and therefore did not investigate the use of different regulatory instruments for managing uncertainty across pre- and post-marketing phases.

Comparative studies of cancer medicines approvals between the EMA and FDA typically focused on the speed with which the two agencies conduct their review and the extent to which the use of conditional approval pathways reduce product development and regulatory review times.[238, 244, 247, 248] While previous research has explicitly compared conditional approval between EMA and FDA,[238] it is unclear how EMA and FDA handle uncertainty in the approval of new cancer medicines, considering the full evidence generation across pre- and post-marketing phases, and a range of regulatory tools for managing uncertainty. Despite the common objective of making promising treatments available to patients faster, and similar regulatory provisions, there does not appear to be consistency in how EMA and FDA have used conditional approval. For example, lapatinib for the treatment of hormone receptor positive metastatic breast cancer received conditional approval by the FDA, but was granted regular approval by the EMA for the same indication. Conversely, vismodegib for treatment of patients with locally advanced basal cell carcinoma received conditional approval by the EMA, but was granted regular approval by the FDA. Such discrepancies have important policy implications since medicines granted regular approval are not subject to the same controls when post-marketing studies are not completed or fail to confirm a positive benefit-risk balance, and may indicate a greater willingness by the regulatory agency to accept uncertainty. Existing comparative studies of FDA and EMA cancer medicines approvals have typically focused on individual aspects of regulatory tools, such as concordance in the use of special vs. regular approval pathways,[238, 244]wording of approved indications, [249] and possible reasons for observed discrepancies in regulatory decisions.[182, 250]

4.1.4. Study aims

This study aimed to investigate how EMA and FDA deal with uncertainty in the clinical evidence about new cancer medicines due to pivotal evidence being based on non-randomised studies and/or the use of surrogate endpoints. The study compared regulatory outcomes (approval vs. no approval decision), pathways (regular approval vs. approval through a special pathway), final approved indications, and regulator-imposed obligations to generate additional evidence in the post-marketing phase for a matched set of cancer medicine indications which received either EMA CMA, FDA AA, or both, indicating that there was less complete data available at time of market entry than usually required for regulatory approval.

4.2. Methods

EMA's and FDA's handling of uncertainty was compared in four steps: first, a set of cancer medicine indications was identified for which regulators judged to have limited evidence on efficacy and safety at the time of approval and still required confirmation. This was operationalised as approval through either the EMA's CMA or FDA's AA pathway, indicating products for which at least one of the two regulators considered the uncertainty about efficacy and safety at the time of market entry substantial enough to preclude regular approval. Second, medicine indications were matched across EMA and FDA. Third, publicly available regulatory documents were reviewed to compile information from EMA and FDA on regulatory pathways and outcomes, final approved indications, pivotal trial characteristics and post-marketing obligations for each indication. Fourth, the compiled information was compared between EMA and FDA. The steps are described in more detail below.

4.2.1. Sample selection

From all cancer medicine indications approved by the EMA (from inception of the CMA pathway in 2006 until 2016) or the FDA (from inception of the AA pathway in 1992 until 2017), approvals through either the EMA CMA or FDA AA pathways were identified from published reports.[82, 156] This list included all cancer medicines for which at least one of the two agencies considered the evidence at the time of approval less complete than required for regular approval. For each of the medicine indications

on this list, targeted database searches on EMA and FDA websites were conducted to identify the closest matching medicine indication with a regulatory outcome by the other agency.

A list of all oncology products approved by the EMA from 2009-2013 inclusive was reported by Davis and colleagues.[85] Additional searches in the EMA database identified oncology medicines (using ATC codes L01-L04) that were denied approval by the EMA in this period or where applications were withdrawn by the manufacturer before a final recommendation was formed (pre-authorisation withdrawals only). Applications for generics, treatments of benign tumours, supportive treatments, and advanced medicine therapy products were excluded.

Matching of FDA decisions with included EMA medicine indications was conducted in a hierarchical fashion. First, potentially matching FDA indications with AA were extracted from Naci and colleagues [61] and Beaver and colleagues.[82] As a minimum, indications had to match on cancer type (e.g. lung cancer, breast cancer, haematologic malignancies), followed by line of treatment, single agent and combination therapy, and population (e.g. type of tumour, biomarkers, histology), if possible. For direct matches with no obvious differences in these parameters, the EMA and FDA decision were declared as match. Next, Drugs@FDA was searched for the non-proprietary names of drugs with EMA CMA to identify matching FDA approved indications based on the same parameters as described above. Finally, the EMA database was searched for the non-proprietary names of drugs with FDA AA to identify potentially matching EMA indications that fell outside the 2009-2013 period covered by Davis et al.[85] Again, indications were declared as direct matches when no obvious differences existed with respect to cancer type, line of treatment, single/combination therapy, and population. When differences existed, both EMA and FDA databases were reviewed again to identify potential better matches (e.g. matching line of treatment was considered a better match than only matching cancer type). If such better matches could be identified through database searches, these were declared as matches. However, if no better match based on the parameters was identified, EMA-FDA indication pairs were still declared as matches as long as the cancer type was the same (i.e. a match could be declared for the same product indicated for first-line treatment by one agency and second-line treatment by the other agency, as long as no regulatory

decision about the same-line treatment was available from both agencies). Differences in final approved wording of indications between matched indication pairs were documented.

An example of a medicine indication for which a better match between EMA and FDA was identified that fell outside the study period was cetuximab for metastatic colorectal cancer in first line combination with irinotecan. This indication was approved through the FDA AA pathway. A similar indication (in combination with FOLFOX) received regular EMA approval in 2012 (as variation of an existing EMA marketing authorisation), thereby making this a potential match with regular approval by one agency and special approval by the other. However, after an additional search of the EMA database, a better match was identified: in 2004, an initial EMA marketing authorisation was granted for cetuximab "in combination with irinotecan in the treatment of patients with epidermal growth factor receptor (EGFR)-expressing metastatic colorectal cancer after failure of irinotecan-including cytotoxic therapy". This better match fell outside the study period and was therefore excluded from this study.

The set of cancer medicine indications was restricted to those with an EMA regulatory outcome (approval, no approval, or application withdrawn by the manufacturer) from 1 January 2009 to 31 December 2013. The time period was chosen to allow for sufficient time for post-marketing obligations to be completed (minimum of five years post-approval until data collection stopped in December 2018).

While the EMA database contains approvals as well as products that were denied approval and marketing authorisation applications withdrawn by the manufacturer after unfavourable assessment by the EMA, the FDA database only provides information on products that were approved. For FDA regulatory decisions, additional web searches were therefore conducted when a medicine was not found in the FDA database.

4.2.2. Data extraction

For each matched medicine indication, information on the regulatory outcome and pathway, current status, final approved indication, pivotal trial(s) at the time of initial

approval, and post-marketing obligations and their current status was compiled from various sources.

Data extraction on regulatory pathways and assessments

Information relating to the regulatory assessment procedure was extracted from publicly available documents at the EMA (European Assessment Reports, EPARs, accessed through the EMA website) and FDA (review documents, label, approval letter, and administrative and correspondence documents, accessed through the Drugs@FDA database) and included information on regulatory outcome and pathway, submission and decision dates, Orphan Drug Designation status, and whether the application was for a first marketing authorisation or a supplemental approval (variation) of a previously approved product. The latter was important information because EMA conditional approval can only be granted for initial authorisations (different from FDA conditional approval, which includes supplemental approvals). Pivotal trial characteristics for most medicine indications were available from two previous studies of all EMA [85] and FDA accelerated approvals.[61] For indications not included in these studies, relevant information on EMA and FDA pivotal trials was extracted from EPARs and review documents and labels, respectively. Where an application for marketing authorisation included more than one indication for the same medicine, these were treated separately in order to assess the evidence submitted and collected in the post-marketing setting for each indication.

Data extraction on post-marketing obligations

Information was also extracted on post-marketing studies that were required by the regulator to confirm the product's efficacy and safety, which are collectively referred to as post-marketing obligations throughout this study. These included specific obligations and obligations ("Annex II conditions") for EMA approvals, and post-marketing requirements under AA ("Subpart E" / "Subpart H") and section 505(o) requirements for FDA approvals. Studies that formed part of routine pharmacovigilance activities, that were conducted voluntarily by the manufacturer, or that were conducted in animals or cells were considered outside the scope of this study.

Studies were categorised as either focusing on confirmation of clinical efficacy and safety or other post-marketing studies (e.g. pharmacokinetics and dosing studies). Only results for the former are reported. Post-marketing obligations were further categorised

according to the study description in the obligation. For example, a study with a primary endpoint of overall or progression-free survival was categorised as efficacy-focused, and a study of adverse events as safety-focused, unless specified otherwise in the obligation.

The post-marketing study population was categorised as either similar or dissimilar to the approved indication. A study population was considered similar to the approved indication when the study was conducted in patients who could receive the drug according to its approved label or when the study was a follow-up of the pivotal trial. A study population was considered dissimilar when the line of treatment was different (e.g. first- instead of approved second-line treatment), when patients had a different cancer (e.g. any solid tumour) or particular type of the same cancer (e.g. a gene mutation), or when other obvious differences existed (e.g. healthy volunteers).

A post-marketing obligation was considered fulfilled when the regulator removed it from the list of required studies, including cases where the manufacturer was released from the obligation without submitting the results from a completed study. For EMA approvals, information on imposed post-marketing studies and their status was extracted from the EPAR, a previous report compiled by the EMA on CMA,[156] and three trial registries (clinicaltrials.gov; EU Clinical Trials Register; and EU PAS Register). In addition, publicly available meeting minutes of two EMA committees responsible for monitoring post-marketing obligations (PRAC and CHMP committees) were reviewed to clarify information provided in other documents. Characteristics of post-marketing studies were extracted primarily from trial register entries, and from the EMA 10-year report on CMA. If no information was available from either of these sources, information included in the EPAR was used. Only post-marketing studies listed in the Annex II to the marketing authorisation under “obligations” (or “specific obligations” in the case of CMAs) were extracted. The status of the obligations was determined primarily through the EMA 10-year report on CMA. For obligations that had not been fulfilled at the time of the cut-off date for the EMA report, information on the current status was obtained from the “steps taken after marketing authorisation” document on the EMA’s website for references to obligations being fulfilled and from historical versions of the Annex II for each product. For dates of fulfilment of obligations, the submission dates of study reports given in the EMA 10-year report on

CMA were used. For obligations that had not been fulfilled at the cut-off date for this report, the "steps taken after marketing authorisation" document and publicly available meeting minutes of the PRAC and CHMP committees were searched for dates when reports on obligations were submitted. When a study report was discussed in one of the meeting minutes, submission of the report was assumed to have been made after the prior meeting of that committee (committees meet monthly). If a submission date could not be identified through these sources, the actual primary completion date or actual completion date (depending on the endpoint of interest of the obligation) of clinical trials register entries was used as proxy for the submission date. Finally, if no clinical trials register entry existed or was not updated, the midpoint between the effective date of the last Annex II version that contained the obligation and the effective date of the first Annex II version where it had been removed served as a proxy for the submission date.

For FDA approvals, information on post-marketing study requirements (i.e. studies subject to reporting requirements) and their status was extracted from the initial approval letter, label changes, the FDA's online database on post-marketing commitments and requirements, and clinicaltrials.gov. Characteristics of post-marketing studies were extracted primarily from clinicaltrials.gov. If no clinicaltrials.gov entry could be identified, information was extracted from the FDA's database. As a minimum, information on required characteristics from approval letters and review documents was extracted. The status of post-marketing studies was determined primarily by perusing all label changes and associated approval letters for the drug in question. The FDA's database of post-marketing commitments and requirements was consulted as secondary data source. Where the status of the requirements could not be determined through approval letters, label changes and the FDA database, study completion status from clinicaltrials.gov were used. For the date of fulfilment of the requirement, submission dates for study reports submitted to the FDA were used as primary source. The FDA's database of post-marketing commitments and requirements was also checked for completion dates. In cases where submission dates of study reports could not be determined from these sources, the actual primary completion date or actual completion date (depending on the endpoint of interest for

the requirement) as stated on clinicaltrials.gov served as proxy for timeliness of study report submission to the FDA.

4.2.3. Analysis

The analysis focused on key features of the regulatory decisions (including regulatory outcome, pathway, pivotal trial evidence, final approved indication, and post-marketing obligations) for each matched medicine indication pair and presents differences and similarities between EMA and FDA.

The different dimensions of regulatory decisions are presented in Table 5. For each combination of regulatory pathway (i.e. same pathway at EMA and FDA, different pathway at one of the agencies, or no approval by one of the agencies), pivotal trial characteristics and, where relevant, post-marketing obligations and final, approved indications were assessed. To illustrate the complex nature of regulatory decision making and do justice to some of the nuances of regulatory outcomes, four case studies that provide additional details on the evidence review and ultimate fate of selected medicines are also presented.

Table 5: Analytical framework for comparing regulatory decisions by EMA and FDA

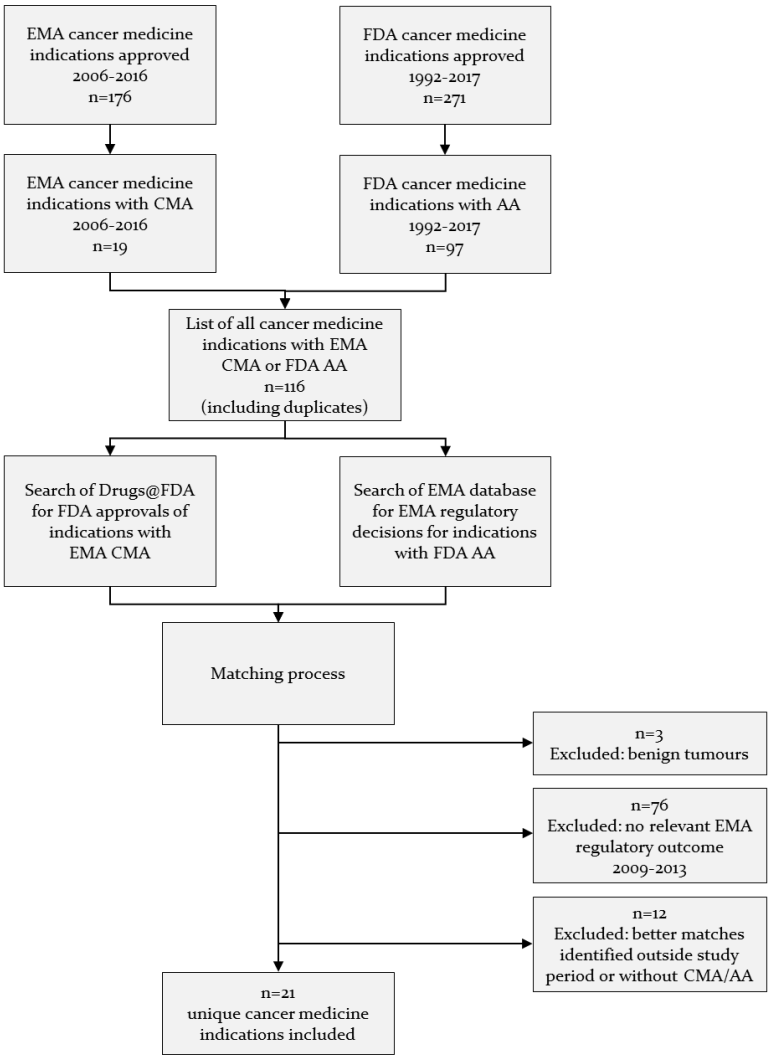
Regulatory pathway for matched medicine indication pairs	**Pivotal trial evidence**	**Post-marketing obligations**	**Approved indication**
Conditional approval by both agencies: • Conditional EMA & FDA approval	Comparison of methodological quality of pivotal trial evidence leading to conditional approval by both agencies, including: • Single-arm trial vs. RCT • Surrogate measure vs. final, patient-relevant outcome	Comparison of methodological and design characteristics of post-marketing obligations imposed by the two agencies, including: • Single-arm trial vs. RCT • Surrogate measure vs. final, patient-relevant outcome • Patient population • Duration of study	Comparison of final, approved indication (narrow vs. broad indication)
Combination of **regular and conditional** approvals: • Regular EMA/conditional FDA approval • Conditional EMA/regular FDA approval	Comparison of the pivotal trial evidence (identical trial vs. different trial) leading to different regulatory outcomes (conditional approval vs. no approval)	Comparison of whether post-marketing obligations were imposed and, if so, their methodological and design characteristics, including: • Single-arm trial vs. RCT • Surrogate measure vs. final, patient-relevant outcome	Comparison of final, approved indication (narrow vs. broad indication)

		• Patient population • Duration of study	
Combination of **conditional approval and no approval**: • No EMA approval/ conditional FDA approval • Conditional EMA approval/ no FDA approval	Comparison of the pivotal trial evidence (identical trial vs. different trial) leading to different regulatory outcomes (conditional approval vs. no approval)	Not applicable (no approval by one of the two agencies)	Not applicable (no approval by one of the two agencies)

4.3. Results

Figure 8 shows the flow chart for selecting eligible cancer medicine indications. From a total of 116 identified indications with limited clinical data on efficacy and safety, as indicated by approval granted through either EMA CMA (n=19; 11% of all approved EMA cancer medicines indications from 2006-2016) or FDA AA (n=97; 36% of all approved FDA cancer medicines indications from 1992-2017) pathways, 21 unique matched indications with EMA CMA or FDA AA were included for which an EMA regulatory outcome between 2009-2013 existed.

Figure 8: Flow chart of selection process for included cancer medicine indications



Abbreviations: AA, accelerated approval; CMA, conditional marketing authorisation.

Table 6 provides an overview of characteristics of the 21 matched cancer medicine indications. Compared to the FDA, the EMA more often granted regular approval (7 [33%] vs. 5 [24%] of matched indications, respectively) or did not grant approval (4 [19%] vs. 1 [5%]), whereas the FDA more often utilised the AA pathway (15 [71%]) compared to the EMA using the CMA pathway (10 [48%]). Overall, there was little overlap in the use of conditional approval pathways at the EMA and FDA. Only four (19%) matched indications received approval through both the EMA CMA and FDA AA pathways. The most common combination of regulatory pathways was regular EMA approval and FDA AA (seven indications [33%]), followed by the reverse, EMA CMA

and regular FDA approval (five indications [24%]). A further four (19%) indications were not approved by the EMA but benefitted from FDA AA, and one indication with EMA CMA approval was not approved by the FDA.

Sixteen of 21 matched indications were first marketing authorisations. The remaining five indications were variations of prior approvals and therefore not eligible for EMA CMA. As of December 2018, half of the indications with EMA CMA had been converted to full (regular) approvals, bringing the total number of regular approved indications to 12 (57%). The other half (n=5) remained under CMA provisions. Conversely, for 13 of 15 indications with FDA AA, all post-marketing obligations had been fulfilled by December 2018, bringing the total number of approvals without pending obligations to 18 (86%). The applications that had not been successful initially (n=4 at the EMA and n=1 at the FDA) remained without approval.

Table 6: Characteristics, regulatory pathways and outcomes for 21 matched cancer medicine indications

	Number of indications (%)	
	EMA	**FDA**
Matched indications, total	**21**	**21**
Regulatory pathway		
Regular approval	7 (33%)	5 (24%)
Special regulatory approval (EMA CMA / FDA AA)	10 (48%)	15 (71%)
No approval	4 (19%)	1 (5%)
Orphan designation	14 (67%)	15 (71%)
First marketing authorisation	16 (76%)	16 (76%)
Regulatory status (December 2018)		
Fully approved	12 (57%)	18 (86%)
Remains under special approval (CMA/AA)	5 (24%)	2 (10%)
Not approved	4 (19%)	1 (5%)
Combinations of regulatory pathways at EMA and FDA		
EMA CMA and FDA AA	4 (19%)	
Regular EMA approval and FDA AA	7 (33%)	
EMA CMA and regular FDA approval	5 (24%)	
No EMA approval and FDA AA	4 (19%)	
No FDA approval and EMA CMA	1 (5%)	

Abbreviations: AA, accelerated approval; CMA, conditional marketing authorisation.

4.3.1. Comparison of regulatory pathways and outcomes, and pivotal trial evidence

Approval pathways, pivotal trial evidence and final approved indications for each of the 21 matched indications were compared between EMA and FDA to further investigate the discrepancies in the use of regulatory pathways between the two agencies and explore possible explanatory factors, such as availability of more robust efficacy and safety data, as well as statutory restrictions on the use of EMA CMA (Figure 9).

Figure 9: Comparison of approval, regulatory pathway, pivotal trial design, and approved indication for matched cancer medicine indications

Generic name	Cancer type	Initial approval	Regulatory pathway	Pivotal trial	Indication
Conditional approval by both EMA (CMA) & FDA (AA)					
brentuximab vedotin	Hodgkin lymphoma [O]	Yes	✓	✓ Same SAT	✗
brentuximab vedotin	Systemic anaplastic large cell lymphoma [O]	Yes	✓	✓ Same SAT	✓
crizotinib	Lung cancer [O – FDA only]	Yes	✓	✗ FDA: 1 additional SAT	✗
ofatumumab	Chronic lymphocytic leukaemia [O]	Yes	✓	✓ Same SAT	✓
Regular EMA approval & FDA AA					
dasatinib	Chronic myelogenous leukaemia [O]	No	✗	✓ Same RCT	✓
imatinib	Gastro-intestinal stromal tumour [O]	No	✗	✓ Same RCT	✗
lapatinib	Breast cancer	No	✗	✓ Same RCT	✓
nilotinib	Chronic myelogenous leukaemia [O – EMA only]	No	✗	✓ Same RCT	✓
pomalidomide	Multiple myeloma [O]	Yes	✗	✗ EMA: RCT FDA: SAT	✗
ponatinib	Chronic myeloid leukaemia [O]	Yes	✗	✓ Same SAT	✗
ponatinib	Acute lymphoblastic leukaemia [O]	Yes	✗	✓ Same SAT	✗
Regular FDA approval & EMA CMA					
bosutinib	Chronic myelogenous leukaemia [O]	Yes	✗	✓ Same SAT	✓
pazopanib	Renal cell carcinoma (first line)	Yes	✗	✓ Same RCT	✓
pazopanib	Renal cell carcinoma (second line)	Yes	✗	✓ Same RCT	✓
vandetanib	Medullary thyroid cancer [O – FDA only]	Yes	✗	✓ Same RCT	✗
vismodegib	Basal cell carcinoma	Yes	✗	✓ Same SAT	✗
No EMA approval & FDA AA					
bevacizumab	Glioblastoma	No	–	✗ FDA: 1 additional SAT	–
omacetaxine mepesuccinate	Chronic myeloid leukaemia [O]	Yes	–	✗ FDA: pooled analysis incl. 1 additional SAT	–
pralatrexate	Peripheral T-cell lymphoma [O]	Yes	–	✓ Same SAT	–
romidepsin	Peripheral T-cell lymphoma [O]	Yes (EMA) No (FDA)	–	✓ Same SAT	–
No FDA approval & EMA CMA					
pixantrone	Non-Hodgkin B-cell Lymphomas	Yes	–	✓ Same RCT	–

Legend: ✓ No discrepancy between EMA & FDA; ✗ Discrepancy between EMA & FDA; – Not approved by one agency

Abbreviations: AA Accelerated Approval; CMA Conditional Marketing Authorisation; [O] Orphan indication; RCT Randomised controlled trial; SAT Single-arm trial

Pivotal trial characteristics

Across the 21 indications, a total of 25 trials provided pivotal evidence, and the majority were single-arm trials (15 of 25 pivotal trials [60%]). Regulatory decisions made by the EMA and FDA were based on the same pivotal trial for 17 of 21 (81%) matched indications. There were only two cases where different trials were submitted to EMA and FDA (pomalidomide and omacetaxine mepesuccinate). For the remaining two pairs, the FDA included a trial as additional "pivotal" evidence while the same trial was only included as "supportive" evidence in the EMA review.

All indications with conditional approval by both EMA (CMA) and FDA (AA) were based on single-arm trial evidence. Among indications with different regulatory pathways across the two agencies, regular EMA approvals with matching FDA AA were more often based on RCTs (5 of 6 pivotal trials [83%]) compared to regular FDA approvals with matching EMA CMA (2 of 4 pivotal trials [50%]).

None of the pivotal trials submitted for 21 marketing authorisation applications at the two agencies used overall survival as a primary endpoint. Indications approved through EMA's CMA and FDA's AA pathways were predominantly based on trials with response rate endpoints, although EMA CMAs also relied on progression-free survival (PFS) or recurrence-free survival in three of 10 trials (30%) compared to two of 17 trials (12%) in FDA approvals with AA. Both EMA (in four cases) and FDA (in two cases) granted regular approval on the basis of trials with response endpoints for matched indications that were approved through conditional pathways (CMA/AA) by the other agency.

Conditional approval by both EMA and FDA

Indications approved through both EMA CMA and FDA AA pathways (n=4) were based on the same single-arm trials across both regulators. In one case (crizotinib), the FDA review included a second pivotal trial which was labelled "supportive" in the EMA review, but was considered an important part of the evidence base also for the EMA's decision.

There were no meaningful differences in the final approved indications in two cases. The EMA used a more restrictive final approved indication for crizotinib, specifying that it was to be used as second- or later-line treatment, whereas the FDA approved treatment of patients with locally advanced or metastatic disease. In the case

of brentuximab vedotin for patients with Hodgkin lymphoma, the EMA restricted its use to patients with CD30-positive disease, mirroring the inclusion criteria of the pivotal trial (see also section 4.3.5 for a detailed description of the approval process for brentuximab vedotin).

Regular EMA and conditional FDA approval

Indications with regular EMA approval and FDA AA (n=7) were in all but one case based on the same pivotal trial (an RCT in four cases and a single-arm trial in two cases). For half (n=3) of these indications, the EMA used more restrictive wording compared to the FDA, whereas no meaningful difference in the final approved indication was detected for the remaining three indications that were based on the same pivotal trial evidence. The one indication where the pivotal trial evidence differed between EMA and FDA was pomalidomide for third-line treatment of multiple myeloma. The basis for the EMA's decision to grant regular approval was an RCT comparing pomalidomide plus low-dose dexamethasone vs. high-dose dexamethasone. This trial had been discussed as a possible pivotal trial for FDA approval but was considered not acceptable by the FDA due to the study design not isolating the treatment effect of pomalidomide. The trial was ultimately accepted by the FDA as the confirmatory trial to satisfy the post-marketing obligations under AA. EMA's regular approval for pomalidomide was restricted to combination therapy with dexamethasone. This restriction was not part of the initial FDA AA but was added upon submission of the results from the confirmatory trial.

Of the seven indications with regular EMA approval and FDA AA, four (dasatinib, imatinib, laptinib, nilotinib) were variations of existing approvals, therefore excluding the possibility of EMA CMA, and three (pomalidomide and two ponatinib indications) were initial marketing authorisation applications, without statutory restrictions regarding the regulatory pathway. The EMA used more restrictive wording for the three initial marketing authorisations, but for only one of the four variations.

Conditional EMA and regular FDA approval

Among indications with EMA CMA and regular FDA approval (n=5), the same pivotal trial evidence was used in all cases (an RCT for three indications and a single-arm trial for two indications), and final approved indications were similar in three cases. For the remaining two indications (vandetanib and vismodegib), both EMA and

FDA used restrictive wording, but it was not possible to determine whether one agency was more restrictive than the other.

No approval

Unsuccessful applications for EMA approval with matching FDA AA (n=4) were in most cases based on the same single-arm trials. For bevacizumab, a second single-arm trial designated as "pivotal" by the FDA was considered "supportive" in the unsuccessful EMA application. For omacetaxine mepesuccinate, the "pivotal trial" submitted to the FDA was a pooled analysis of two single-arm trials, both of which were part of the evidence package submitted to the EMA (one as pivotal and one as a supporting trial). Three of the unsuccessful EMA applications were for initial marketing authorisations (omacetaxine mepesuccinate, pralatrexate, and romidepsin), where CMA was also possible.

Finally, in the case of the unsuccessful FDA application which received EMA CMA (pixantrone), the same RCT was used as pivotal trial evidence.

4.3.2. Comparison of post-marketing obligations

There was overall considerable heterogeneity between EMA and FDA with respect to the number of post-marketing studies required, their objectives, and design characteristics. Key characteristics of clinical post-marketing obligations at the aggregate level are presented in Table 7, and study-level details are presented in Table 8. A comparison of post-marketing obligations at the indication level is shown in Figures 10-12.

Overall, the EMA imposed clinical post-marketing obligations (i.e. obligations for confirmatory studies of clinical efficacy and/or safety in patients) for 12 of 17 approved indications (71%) and the FDA for 17 of 20 approved indications (85%). The mean number of clinical post-marketing obligations was 2.3 (range, 1 to 4) for EMA CMAs and 0.6 (range, 0 to 3) for EMA regular approvals, and it was 1.8 (range, 1 to 4) for indications approved through the FDA AA pathway and 0.4 (range, 0 to 1) for FDA regular approvals. There was only small overlap between EMA and FDA in individual studies used to address post-marketing obligations. Post-marketing obligations were addressed by a total of 43 unique clinical studies, but only two of these were used for both EMA and FDA obligations.

Key characteristics of clinical post-marketing obligations are shown in Table 7. Post-marketing obligations included at least one RCT in six of all 12 EMA approvals with obligations (50%) compared to 14 of all 17 FDA approvals with obligations (82%). EMA approvals included single-arm or observational studies for nine of 12 approvals (75%), whereas only five of 17 (29%) approved FDA indications with clinical confirmatory studies included these non-comparative study designs. Overall survival was a primary endpoint in post-marketing studies for only three of 12 approved EMA indications (25%) and four of 17 (24%) FDA indications. These studies were RCTs in one EMA and three FDA approvals.

Across both agencies, the majority of post-marketing obligations for products approved through the CMA or AA pathways was conducted in populations similar to the approved indication, i.e. patients falling under the approved label or follow-up studies of the pivotal trial. However, while 87% of EMA post-marketing obligations were conducted in similar populations, this was only the case for 59% of FDA obligations attached to AA, with the remaining 41% conducted in different populations, such as previously untreated patients or patients with different cancer types (Table 8).

4.3.3. Status of post-marketing obligations

The status of post-marketing obligations was assessed as of December 2018, resulting in a median follow-up time of 7.25 years overall (median 6.6 years for EMA and 7.5 years for FDA approvals). After this time, half of EMA approvals with clinical post-marketing obligations had all requirements fulfilled. However, there were delays for five of these and all obligations had been fulfilled on time for only one approved indication. The remaining 50% of approved EMA indications had open obligations, which were all running behind schedule. Of the 17 approved FDA indications with clinical post-marketing obligations, five (29%) had fulfilled all obligations on time, and obligations had been fulfilled with a delay for eight indications (47%). Two approved FDA indications had open obligations running on time, and two with delays. EMA obligations had a median delay of 10 months (range, no delay to 5 years), and FDA obligations had a median delay of 0 months (range, no delay to 3 years), although this included five cases where the manufacturer was released from an obligation and this was therefore considered fulfilled on time by the FDA (one obligation each for

brentuximab vedotin for systemic anaplastic large cell lymphoma, crizotinib, lapatinib, omacetaxine mepesuccinate, pomalidomide). There were shorter delays for obligations for indications approved under EMA CMA or FDA AA (EMA: median 6 months; FDA: median 0 months) compared to regular approvals (EMA: median 29 months; FDA: median 17.5 months).

Table 7: Key characteristics of clinical post-marketing obligations (PMOs) imposed by EMA and FDA

	EMA			FDA		
	CMA	**Regular approval**	**Overall**	**AA**	**Regular approval**	**Overall**
Number of indications with clinical PMOs	**10**	**2**	**12**	**15**	**2**	**17**
Post-marketing obligation study design						
Included at least one RCT	5	1	6 (50%)	13	1	14 (82%)
Included at least one single-arm trial	4	1	5 (42%)	3	0	3 (17%)
Included at least one observational Phase IV	5	1	6 (50%)	1	1	2 (12%)
Post-marketing obligation study endpoint						
Included at least one study with overall survival as primary endpoint	2	1	3 (25%)	3	1	4 (24%)
Included at least one RCT with overall survival as primary endpoint	0	1	1 (8%)	2	1	3 (18%)
Post-marketing obligation status (December 2018)						
All obligations fulfilled on time	1	0	1 (8%)	5	0	5 (29%)
All obligations fulfilled, with delays	4	1	5 (42%)	7	1	8 (47%)
Has open obligations, running on time	0	0	0 (0%)	1	1	2 (12%)
Has open obligations, running with delays	5	1	6 (50%)	2	0	2 (12%)

Abbreviations: AA, accelerated approval; CMA, conditional marketing authorisation; PMO, post-marketing obligation.

Table 8: Characteristics of clinical post-marketing study obligations

	EMA		FDA	
	CMA	**Regular approval**	**AA**	**Regular approval**
	N (%)	**N (%)**	**N (%)**	**N (%)**
Total number of obligations	**23**	**4**	**27**	**2**
Required under CMA/AA	22 (96%)	0 (0%)	21 (78%)	0 (0%)
Other obligations	1 (4%)	4 (100%)	6 (22%)	2 (100%)
Focus of post-marketing study				
Focus on efficacy and safety	10 (43%)	1 (25%)	7 (26%)	0 (0%)
Focus on efficacy only	10 (43%)	2 (50%)	15 (56%)	0 (0%)
Focus on safety only	3 (13%)	1 (25%)	5 (19%)	2 (100%)
Post-marketing study design				
RCT	8 (35%)	2 (50%)	22 (81%)	1 (50%)
Single-arm trial	9 (39%)	1 (25%)	4 (15%)	0 (0%)
Observational Phase IV	6 (26%)	1 (25%)	1 (4%)	1 (50%)
Active control arm	8 (35%)	1 (25%)	13 (48%)	0 (0%)
Placebo control arm	0 (0%)	1 (25%)	8 (30%)	1 (50%)
Double blind	0 (0%)	1 (25%)	4 (15%)	1 (50%)
Open label	22 (96%)	3 (75%)	21 (78%)	1 (50%)
Post-marketing study endpoint				
Overall survival	3 (13%)	1 (25%)	4 (15%)	1 (50%)
Progression-free survival	11 (48%)	1 (25%)	11 (41%)	0 (0%)
Adverse events	6 (26%)	2 (50%)	3 (11%)	0 (0%)
Response rate	7 (30%)	1 (25%)	5 (19%)	0 (0%)
Other surrogate endpoint	0 (0%)	1 (25%)	4 (15%)	0 (0%)
Unclear endpoint	0 (0%)	0 (0%)	1 (4%)	1 (50%)

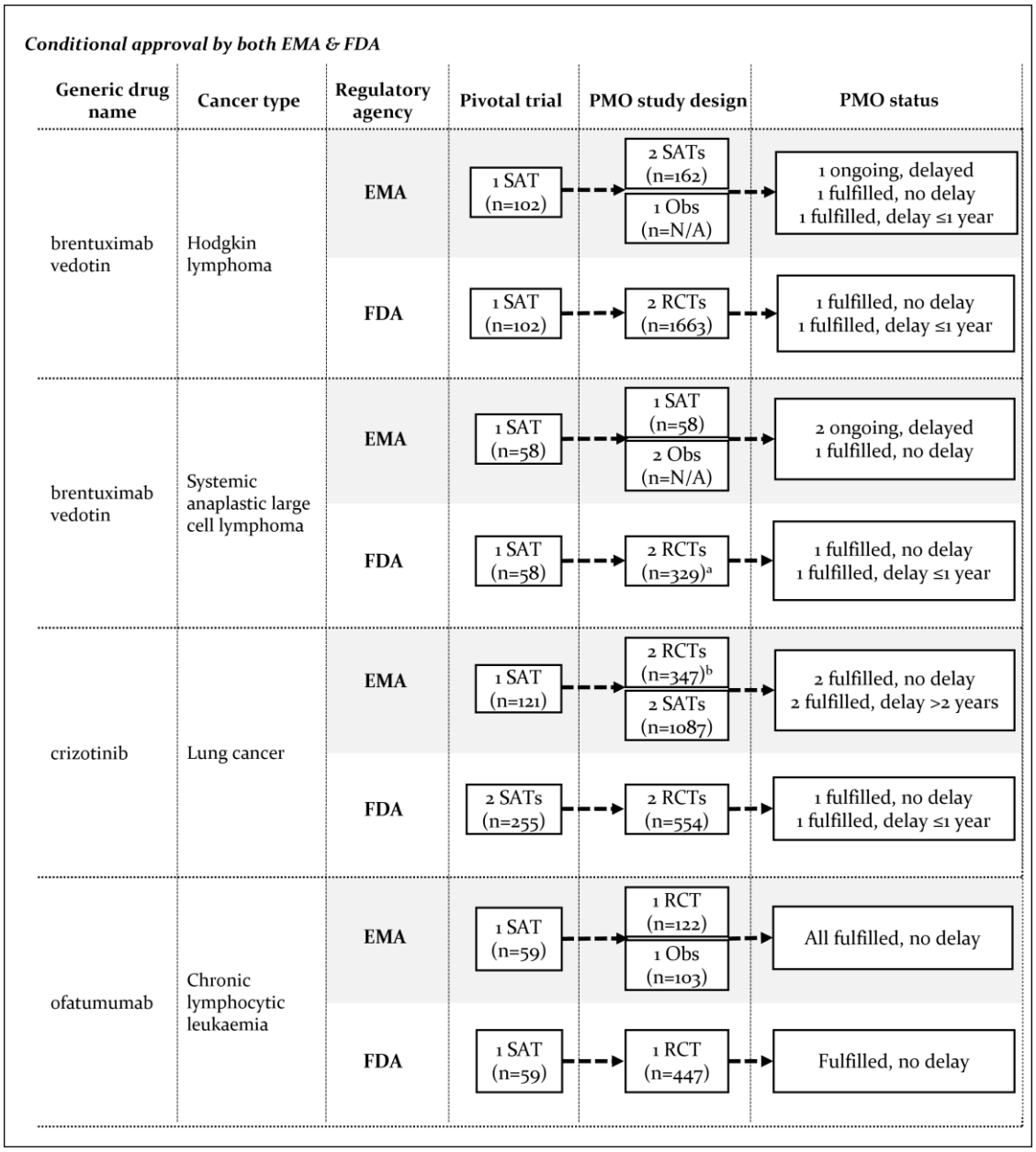
Post-marketing study population				
Study population similar to approved indication	20 (87%)	2 (50%)	16 (59%)	1 (50%)
Study population different to approved indication	3 (13%)	2 (50%)	11 (41%)	1 (50%)
Status (December 2018)				
Ongoing, on track	0 (0%)	0 (0%)	1 (4%)	1 (50%)
Ongoing, delayed	6 (26%)	1 (25%)	2 (7%)	0 (0%)
Fulfilled, on time	9 (39%)	1 (25%)	13 (48%)	0 (0%)
Fulfilled, delayed by ≤1 year	6 (26%)	0 (0%)	6 (22%)	0 (17%)
Fulfilled, delayed by ≤2 years	0 (0%)	1 (25%)	3 (11%)	0 (0%)
Fulfilled, delayed by ≤3 years	0 (0%)	1 (25%)	0 (0%)	1 (50%)
Fulfilled, delayed by ≤4 years	2 (9%)	0 (0%)	0 (0%)	0 (0%)
Status unclear	0 (0%)	0 (0%)	2 (7%)	0 (0%)

Abbreviations: AA, accelerated approval; CMA, conditional marketing authorisation; PMO, post-marketing obligation.

4.3.4. Post-marketing evidence generation

Among medicine indications that received approvals through both EMA CMA and FDA AA pathways (n=4), EMA obligations included a total of nine non-randomised studies (comprising follow-up of pivotal single-arm trials, new single-arm trials, observational safety studies) and a total of three RCTs (Figure 10). The RCTs were requested for only two of the four indications (two RCTs for crizotinib and one RCT for ofatumumab), with no RCT imposed for the remaining two indications (both brentuximab vedotin). In contrast, FDA obligations included two RCTs each for three of the four indications, and one RCT in the remaining indication. The FDA considered the evidence generation obligations under AA for brentuximab vedotin fulfilled on the basis of the results from one of the required RCTs, while the medicine remained under CMA provisions in Europe, with confirmatory studies ongoing and delayed by 2 to 5 years (see also section 4.3.5 for details on the regulatory review process and follow-up for brentuximab vedotin). For crizotinib, the same RCT was used in both Europe and the US to confirm clinical benefit. Ofatumumab was also converted to regular approval by both agencies, although on the basis of different studies: the confirmatory study requested by the EMA was a new RCT conducted in patients with refractory disease (approved indication),[251] whereas the confirmatory FDA RCT was conducted in the first-line setting.[252]

Figure 10: Pre-marketing and post-marketing evidence requirements and status of clinical post-marketing obligations for matched indications with EMA CMA and FDA AA

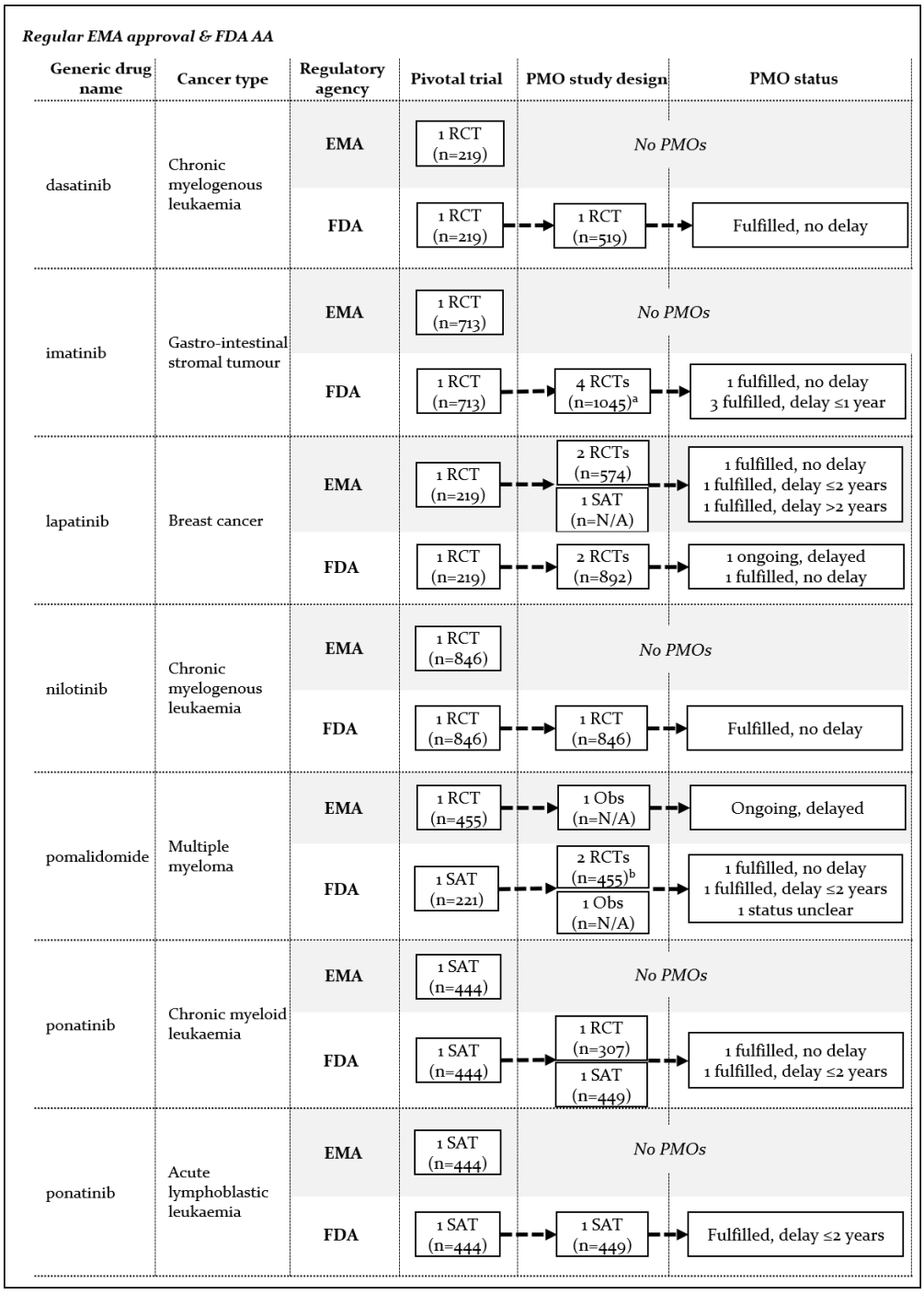


Abbreviations: N/A, sample size not available because study is ongoing; Obs, observational phase IV study; PMO, post-marketing obligation; RCT, randomised controlled trial; SAT, single-arm trial.
Footnotes: [a] Sample size is for one RCT only, as the obligation for the other RCT was considered fulfilled through a related obligation in the Hodgkin lymphoma indication.
[b] Two obligations relating to the same RCT.

Among indications approved through regular EMA and FDA AA pathways (n=7), post-marketing studies were imposed by the EMA for only two indications (lapatinib and pomalidomide), while the FDA imposed obligations for all indications (Figure 11).

FDA obligations included a minimum of one RCT per indication, with the exception of ponatinib for acute lymphoblastic leukaemia (see section 4.3.5 for details). By December 2018, the FDA considered post-marketing obligations fulfilled for all except one medicine (lapatinib), with confirmatory evidence submitted to the FDA on time in two cases (dasatinib and nilotinib) and with delays ranging between 11 and 20 months for the remaining indications. In the case of lapatinib, the endpoint for the confirmatory trial was changed from overall survival (as originally indicated) to PFS. For a second post-marketing RCT for this product, the primary endpoint was also changed from overall survival to PFS, with results indicating inferiority of lapatinib compared to trastuzumab for this endpoint, and no statistically significant difference for mortality (although more deaths were reported in the lapatinib arm compared to trastuzumab).[253] However, there was no communication regarding the submission of the results of this trial in FDA documents.

Figure 11: Pre-marketing and post-marketing evidence requirements and status of clinical post-marketing obligations for matched indications with regular EMA approval and FDA AA



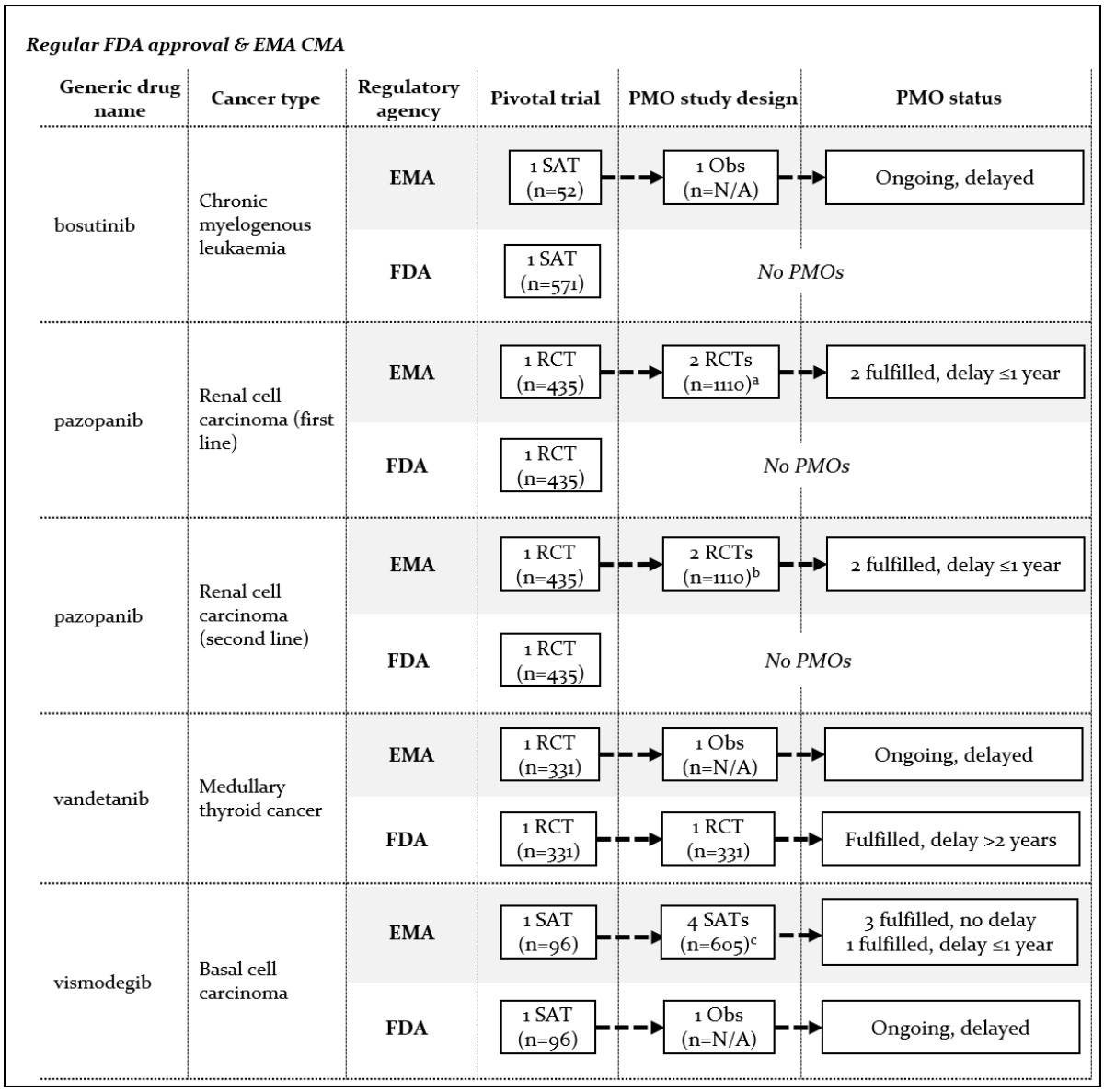
Abbreviations: AA, FDA Accelerated Approval; N/A, sample size not available because study is ongoing; Obs, observational phase IV study; PMO, post-marketing obligation; RCT, randomised controlled trial; SAT, single-arm trial.

Footnotes: [a] Four obligations relating to two RCTs.

[b] Sample size is for one RCT only. Obligation for the other RCT was considered fulfilled through another study.

Among indications with EMA CMA and regular FDA approval (n=5), the EMA imposed post-marketing studies for all conditionally approved indications, while the FDA only imposed obligations for two of them (vandetanib and vismodegib) (Figure 12). EMA obligations included a RCT for only two of five conditionally approved indications. Confirmatory studies for the remaining conditionally approved indications included observational phase IV studies, a follow-up of the pivotal single-arm trial, and a new single-arm trial. Pazopanib and vismodegib were converted to regular approval after delayed (six and 11 months, respectively) submission of confirmatory study reports. Bosutinib and vandetanib remained under CMA, with submission of the confirmatory study reports having been extended, respectively, by three years and eight months, and four years and nine months due to slow recruitment.[156, 254]

Figure 12: Pre-marketing and post-marketing evidence requirements and status of clinical post-marketing obligations for matched indications with regular FDA approval and EMA CMA



Abbreviations: CMA, EMA conditional marketing authorisation; N/A, sample size not available because study is ongoing; Obs, observational phase IV study; PMO, post-marketing obligation; RCT, randomised controlled trial; SAT, single-arm trial.
Footnotes: [a, b] Two obligations relating to the same RCT.
[c] Four obligations (including 1 pooled analysis) relating to two SATs.

4.3.5. Case studies of four cancer medicines with uncertain therapeutic value

The sub-sections below present case studies of the approval of four medicines (for six indications), representing one case each of conditional approval granted by both agencies (brentuximab vedotin), regular EMA and conditional FDA approval (ponatinib), regular FDA and conditional EMA approval (bosutinib), and no EMA and conditional FDA approval (pralatrexate).

Conditional approval by both EMA and FDA: brentuximab vedotin

The manufacturer sought approval for brentuximab vedotin in two separate indications, treatment of Hodgkin lymphoma (HL) and systemic anaplastic large cell lymphoma (sALCL). Joint applications for the two distinct indications were submitted to FDA first and three months later to EMA. The same single-arm, phase II pivotal trials were submitted to EMA and FDA to support the application for the two indications (SG035-0003 for HL and SG035-0004 for sALCL), although the FDA preferred a secondary endpoint (complete remission rate) as regulatory efficacy endpoint for the approval in HL.

The EMA considered the criteria for CMA fulfilled, including established positive benefit-risk balance, likely provision of additional clinical data, unmet medical need, and benefits of immediate availability of the product outweighing the risks associated with the uncertainties in available evidence. Despite the application relying on single-arm trials, the EPAR explicitly mentions PFS as indicative of clinical benefit: "In the patients with relapsed or refractory CD30+ Hodgkin lymphoma following autologous stem cell transplant or patients with relapsed or refractory systemic anaplastic large cell lymphoma brentuximab vedotin showed efficacy in terms of a significant increase in ORR [overall response rate] and PFS, despite the absence of confirmatory controlled data." [88, p98]

While the manufacturer had applied for regular approval in the US, the FDA had advised prior to submission that based on the evidence being prepared (single-arm trials with response rates as endpoints), conditional approval was appropriate, and ultimately did not grant regular approval. Specifically, concerns were raised in relation to the limited safety database. The FDA also mentioned impossible attribution of

adverse events and evaluation of time-to-event endpoints due to the single-arm study design.[89]

The final approved indication for HL was more restrictive in Europe compared to the US, closely mirroring the patients included in the pivotal trial:

> *"The inclusion criteria of both pivotal trials specified that only patients with proven CD30-positive malignancies were allowed. While sALCL is always CD30-positive, a subset of Hodgkin's lymphomas is CD30 negative. It is therefore appropriate that the indication in the treatment of patients with relapsed or refractory Hodgkin's lymphoma (HL) only includes patients with CD30-positive HL."[88, p97]*

In contrast, the approved indication by the FDA was broadened during labelling discussions to include patients in whom autologous stem cell transplant was not an option, despite these patients not being having been included in the pivotal trial.[89] This group was also included in the EMA approved indication.

EMA post-marketing obligations included overall survival follow-up of the pivotal trials, additional single-arm trials (including one in patients not suitable for ASCT), and an observational post-authorisation safety study. FDA obligations, on the other hand, included one confirmatory efficacy RCT for each indication (PFS as endpoint) and one safety RCT. While EMA approval remained conditional as of December 2018, with the new single-arm trial in sALCL patients and the post-authorisation safety study running behind with delays of five and two years, respectively, the FDA accepted submission of the results from the confirmatory RCT in HL patients to convert to regular approval for both indications and released the manufacturer from the requirement to perform a RCT in sALCL patients. The confirmatory trial demonstrated superiority of brentuximab in combination with a chemotherapy regimen versus chemotherapy as frontline therapy in 2-year modified PFS, but not overall survival.[255]

The case study of brentuximab vedotin demonstrates that regulators (in this case, the FDA) may react to a limited evidence package by granting conditional approval, even though the applicant may consider the evidence sufficient for regular approval. It further shows that similar evidence may be interpreted differently by regulators, as evidenced by differences in the final approved indication. Finally, despite lack of

demonstrated overall survival benefit, conditional approval was converted into regular approval in the US due to a demonstrated benefit on PFS, a surrogate endpoint.

Regular EMA and conditional FDA approval: ponatinib

The manufacturer sought approval for two separate indications for ponatinib: treatment of chronic phase, accelerated phase, or blast phase chronic myeloid leukaemia (CML) and treatment of Philadelphia chromosome positive acute lymphoblastic leukaemia (Ph+ ALL). Both EMA and FDA applications were based on the same single-arm pivotal trial (PACE). The EMA considered the pivotal trial to have demonstrated clinically meaningful responses, because "[u]ltimately, response to therapy may enable certain patients to proceed to allogeneic stem cell transplantation, a potentially curative intervention in CML"[256, p97] and pre-specified endpoints for Ph+ patients were achieved. The FDA review team also considered a positive benefit-risk balance established: "The response rates demonstrated in the PACE trial were higher than drugs that we have previously approved for CML or Ph+ALL."[257, p29] However, the FDA had previously required longer follow-up data (24 months) to grant regular approval or convert from conditional to regular approval for other drugs used in CML patients. Granting conditional approval for ponatinib on the basis of the PACE trial with a median of 10 months follow-up was therefore consistent with prior FDA actions. The FDA review team did not consult its standing advisory committee but did request input from two clinical experts. Both stated that the endpoints were reasonably likely to predict clinical benefit and recommended accelerated approval, although one raised concerns over the evidence threshold that was applied for this class of drugs: "[…] I am not entirely comfortable with the endpoint of hematologic improvement in BP-CML and ALL translating to clinical benefit. I worry that we are lowering the bar for TKIs [tyrokinase inhibitors], as such a surrogate would not be accepted for acute myeloid leukemia, for which the survival is similar, if not slightly better."[257, p4]

The FDA raised concerns over a high rate of serious cardiovascular and thromboembolic adverse events in the pivotal trial, which, due to the single-arm design, could not be attributed clearly to ponatinib (the manufacturer claimed that patients were already at high risk of cardiovascular events). These concerns were not discussed with the same urgency in the EMA report, which concluded that risks appeared manageable. In response to its concerns, the FDA decided to include a black

boxed warning in the label and required additional safety data from a RCT comparing ponatinib to nilotinib in the first-line CML setting (EPIC trial). This trial was terminated early due to increased frequency of safety events in other ponatinib trials (follow-ups of the pivotal phase II trial, as well as a phase I trial). At the time of termination of EPIC, there were significantly more serious arterial occlusive events in the ponatinib arm compared to the nilotinib arm,[258] and the ponatinib FDA label was amended to state that the drug was not recommended in patients with newly diagnosed CML. Because of the safety signal in the follow-up of the pivotal phase II trial and the phase I trial, ponatinib was temporarily withdrawn from the US market in 2013, but reintroduced only seven weeks later with a narrowed indication.[259] Prompted by the FDA action, the EMA also reviewed the evidence and concluded that the benefit-risk balance remained positive, taking into consideration updates to the product information to reflect new data on safety.[260]

In the US, ponatinib approval was ultimately converted to regular in November 2016 on the basis of four years of follow-up data from the pivotal PACE trial.[261] Initially, only two years of follow-up had been requested, but the conversion was only made after data from additional two years of follow-up were submitted.

The ponatinib case demonstrates scope for considerable difference in interpretation of safety data from the same clinical trial. Importantly, such differences can result in the availability or lack of important post-marketing evidence: a RCT providing additional safety data was requested by the FDA due to safety concerns identified; the EMA, on the other hand, did not request such additional data, but did review the RCT data as it became available. For the US, the ponatinib approval also shows that prior decisions can provide an important reference point for regulators, as they aim for consistency in regulatory outcomes; in this case, conditional approval was considered consistent with prior decisions based on similar data.

Regular FDA and conditional EMA approval: bosutinib

Bosutinib for previously treated CML received conditional approval by the EMA and regular approval by the FDA. Applications to both agencies relied on the same pivotal single-arm trial, which was not originally intended to be the pivotal trial. The manufacturer had planned to seek approval for first-line CML treatment on the basis of a phase III RCT comparing bosutinib to imatinib. However, the trial failed to

demonstrate superiority of bosutinib and became a supporting trial for the revised indication in previously treated patients both at EMA and FDA.

The EMA proposed conditional approval to the manufacturer on the basis that a positive benefit-risk balance had been established in a subpopulation of patients with chronic, advanced, and blast phase CML with unmet medical need, where bosutinib would be the only treatment option after prior treatment with one or more TKIs and when imatinib, nilotinib and dasatinib are not appropriate. However, additional data was considered necessary, which would come from a planned confirmatory single-arm trial with approximately 150 patients in the same patient population as the approved indication, which was deemed feasible by the EMA CHMP. "Such a study is considered appropriate to allow a conditional marketing authorisation and to confirm the preliminary results observed in the target population as approved by CHMP."[262, p85]

In the US, the manufacturer had planned to seek accelerated approval for the originally planned indication in the first-line setting.[263] Approval in the revised indication was regular because similar evidence (study design, endpoint, and follow-up) had been used by the FDA for conversion of accelerated to regular approval for other drugs in this setting.[264] However, there was substantial disagreement in the assessment of the evidence between the primary statistical reviewer, who did not consider the evidence sufficient to support approval, and the rest of the review team, who, based on the size of the response rate and duration of responses, considered efficacy to be established.[265] In particular, the statistical reviewer considered the evidence in patients in advanced phases of the disease exploratory (the primary endpoint was only specified for patients in chronic phase),[266] but this assessment was overruled by the cross-discipline team leader and office director.

The confirmatory single-arm trial requested by the EMA was due in September 2018. In its October 2018 meeting, the CHMP extended the due date to May 2022,[254] which would constitute a delay of three years and eight months compared to the original due date. EMA approval therefore remained conditional as of December 2018. The FDA requirement for a pharmacokinetics trial was completed on time.

The case study of bosutinib reveals the potential use of conditional approval as regulatory tool to allow, rather than deny approval, as evidenced by the proposal of the EMA to the applicant to seek CMA. Similar to the ponatinib approval described above,

it also highlights the importance of prior regulatory outcomes in the FDA's reasoning for granting regular vs. conditional approval.

No EMA approval and conditional FDA approval: pralatrexate

Pralatrexate was granted conditional approval by the FDA for treatment of patients with relapsed or refractory peripheral T-cell lymphoma (PTCL) in 2009, but the application for EMA CMA for second- or higher-line treatment of PTCL (submitted in 2010) was not successful.

Both applications were based on the same pivotal single-arm trial (PDX-008). However, EMA and FDA had differing opinions on the study design and choice of endpoint. The pivotal trial was designed under FDA special protocol assessment agreement and therefore considered acceptable by the review team: "Because of the rarity of PTCL and the absence of effective therapies for patients with relapsed or refractory PCTCL it was agreed that depending on the magnitude of the response rate, the duration of response, and the risk benefit ratio, a single study in at least 100 patients may be sufficient to support approval."[267, p2] The EMA, on the other hand, did not consider benefit established due to the clinical trial design, which was non-comparative and used response rate as an endpoint, both against prior advice of the EMA CHMP to the applicant. The EPAR states with respect to the submitted trial that "the clinical benefit of objective response per se has not been established in this clinical setting and in the absence of dramatic activity, the clinical benefit cannot be considered established. Additionally, a single arm design does not allow estimation of clinical benefit in terms of clinical benefit endpoints such as PFS or OS."[268, p62]

Both regulators sought input from clinical experts through advisory panels. While EMA advisors stated that "[t]he antitumor activity observed for pralatrexate cannot be considered impressive or otherwise outstanding" [268, p62] and did not consider clinical benefit to be established, the FDA advisory committee "felt that the results of the study were reasonably likely to predict for clinical benefit, such as an improvement in overall survival or a robust improvement in progression-free survival."[267, p12]

The EMA did not consider a positive benefit-risk balance established and denied market approval, whereas the FDA granted accelerated approval because of a lack of alternative therapies in this patient group, a response rate in the pivotal trial that was considered relevant in patients that were heavily pre-treated, and the opinion of the

advisory committee that response rate and duration were reasonably likely to predict clinical benefit.[267] Within the EMA CHMP, there was a divergent minority opinion that considered conditional approval appropriate, based on a response rate that was similar to that observed for other approved agents, unmet medical need, and a proposed confirmatory trial (PDX-3501) that was considered feasible. The majority opinion of the CHMP was also re-examined on the manufacturer's request, but confirmed its negative recommendation.

Post-marketing obligations imposed by the FDA included two confirmatory RCTs, neither of which have been completed. The first was terminated in 2014 after only 18 patients had been enrolled. The second RCT was due to be submitted in September 2015, but there was no evidence of the FDA considering this requirement fulfilled, indicating that the obligation is running behind by over three years.

The case of pralatrexate reveals limits to the acceptable uncertainty in clinical evidence at the EMA, although not the FDA. The same evidence was considered insufficient for approval by the EMA, but was accepted for conditional approval by the FDA. Interestingly, the study design had been subject to scientific advice from both agencies. The EMA had advised for a more robust study design measuring a different outcome and, since this was not provided, denied approval. In contrast to other applications based on limited evidence, conditional approval was not granted because a positive benefit-risk ratio was not deemed established due to the methodological weaknesses of the pivotal trial.

4.4. Discussion

4.4.1. Summary of main findings

This study analysed how EMA and FDA managed uncertainty in the clinical evidence due to the use of non-randomised studies and/or surrogate endpoints for regulatory approval of new cancer medicines. Reviewing regulatory outcomes, pathways, pivotal trial evidence, final approved indications, and post-marketing obligations of 21 matched cancer medicines indications with uncertainties in the pivotal evidence, the study identified differences in regulatory decisions between EMA and FDA. Both agencies showed an overall high acceptance of uncertainty, with the vast

majority of applications with limited evidence at the time of market entry being approved by both regulators. Single-arm trials provided pivotal evidence for 13 of the 21 indications. Although both agencies overwhelmingly relied on the same evidence base, there were frequent discrepancies in the use of conditional (EMA CMA and FDA AA) vs. regular approval pathways across the two settings. Finally, there were marked differences in the design and objectives of post-marketing obligations imposed by EMA and FDA, with EMA obligations more commonly featuring non-randomised studies, including single-arm trials and observational studies.

4.4.2. Discrepancies in the use of conditional approval pathways

While both EMA and FDA were more likely to grant, rather than withhold, marketing approval for the cancer medicine applications in this cohort of matched indications, there was little concordance in the use of regulatory pathways. There were only four of 21 cases where both EMA and FDA granted approval through CMA and AA pathways, respectively, for the same indication. Among 16 medicine indications with approvals by both agencies, there were 12 that obtained approval through one of these conditional regulatory pathways, but received regular approval by the other agency.

Comparative analysis of these cases highlights that the use of conditional vs. regular approval pathways is an important determinant for the generation of a full evidence package. Additional evidence on efficacy and safety typically only became available through post-marketing obligations imposed as part of conditional approval (there were only four cases where such evidence was requested by the agency granting regular approval). The case of ponatinib is helpful to illustrate the role of mandating the collection of post-marketing evidence (see section 4.3.5 for a brief case study). While the EMA granted regular approval on the basis of a single-arm trial, the FDA granted approval through the AA pathway, in part due to a potential safety issue detected in that same single-arm trial. The RCT requested by the FDA as condition for AA substantiated these safety concerns.[258] Both EMA and FDA took this additional evidence in the post-marketing setting into consideration.[256, 259] Importantly, a post-marketing assessment of the benefit-risk profile based on robust evidence was only possible because additional evidence had been collected under the FDA AA provisions.

Previous research has shown that the majority of clinical trials in the post-marketing setting conducted for FDA-approved products relates to new or extended indications of previously authorised treatments.[269] The results presented in section 4.3.2 show that post-marketing obligations were more likely to be conducted in indications similar to the initially authorised indication when conditional approval was granted. Taken together, these findings underline the importance of regulatory standards for initial marketing authorisation and for generating post-marketing evidence: unless regulators demand robust evidence for a specific indication at the time of marketing authorisation, such evidence is unlikely to materialise in the post-marketing setting.

There are several possible reasons for discrepancies in the use of conditional vs. regular approval by EMA and FDA. First and foremost, the EMA can only grant CMA for first (or initial) marketing authorisations, but not for any subsequent changes, such as extensions or variations of existing marketing authorisations. In these situations, the EMA faces a decision to grant approval (if the benefit-risk balance is deemed positive) or not (if it is negative). For the products included in this study, the EMA appears to follow a more lenient approach in these cases: approval was granted in four cases, whereas no approval was given in only one of the cases where the statutory limitation on CMA applied. A more restrictive indication wording, which can be regarded as another tool to manage uncertainty, was used by the EMA in only one of the approvals.

Another possible explanation for the observed discrepancy in granting CMA/AA vs. regular approval is that EMA and FDA based their assessments on different evidence. Another study of concordance between EMA and FDA across all therapeutic areas from 2014 to 2016 found overall very good agreement between the two agencies for approval vs. no approval decisions, but, similar to this study, considerable disagreement in the use of special approval pathways.[182] In their study, Kashoki et al. found that discordance in the use of regulatory approval pathways was in equal parts due to differences in the conclusions on efficacy drawn from the same clinical data submitted to both regulators, as well as additional clinical data reviewed by one agency but not the other. This was different for the present study, which focused on cancer medicines with limited data on efficacy and safety at the time of regulatory decision making. There were only two cases in this study where different pivotal trials were submitted to

EMA and FDA – all others relied on the same clinical trial evidence but resulted in different regulatory outcomes. Moreover, for 13 of the 19 indications based on the same pivotal trials, the results of the trials used for assessment were exactly the same. There was only one case (vandetanib) where more favourable results due to different analysis of the same trial could explain a different regulatory outcome (i.e. the FDA analysis showing greater improvement in PFS as compared to the analysis relied on by the EMA, with the FDA granting regular approval while the EMA granting CMA).

EMA and FDA therefore do not appear to use CMA and AA approval congruently, a finding which confirms and extends what other researchers have reported. For example, Martinalbo and colleagues showed that, of all 16 cancer medicines approved from 2013 to 2014 under special approval pathways, only one was approved through both EMA CMA and FDA AA.[244] Hoekman et al. found that five of 11 cancer medicines with EMA CMA were also approved through the FDA AA pathway, while five received regular approval and one was not approved.[238] In a qualitative study with representatives from EMA and FDA, Tafuri et al. reported that respondents attributed divergent opinions on cancer medicines to the fact that EMA regarded PFS as clinical benefit in itself, whereas FDA saw it as a surrogate endpoint which would need to be confirmed by additional studies.[250] Such discrepant views could explain a situation where the same evidence package leads to the FDA utilising the AA pathway and the EMA granting regular approval. However, in the present study, only one case existed where different views on PFS could have explained approval through the regular EMA and FDA AA pathways (lapatinib).

4.4.3. Evidence standards for regular approval

An important finding of this study is that, despite discrepancies in the use of EMA CMA and FDA AA, there was an overall high acceptance of uncertainty for the approval of cancer medicines by both EMA and FDA. The sample consisted exclusively of indications where at least one of the two regulatory agencies considered the evidence base insufficient to grant regular, or full, approval. Nevertheless, for more than half of the 21 matched indications, one of the agencies did grant regular approval, indicating that regulators often did not consider it necessary to impose obligations to collect additional evidence under the framework of CMA or AA. In addition, regulators were

far more likely to grant approval than to deny it: there were only five indications where one of the agencies did not grant approval (four of the negative decisions coming from the EMA and only one from the FDA).

Conditional approval was intended as exceptional pathway to grant early access to new treatments based on limited data on efficacy and safety. Granting regular approval based on the same data, as shown for a substantial proportion of products in this study, suggests that the exception is becoming the norm. The trajectory of exceptional standards applied for cancer medicines approval, also documented by others,[55, 270] may signal to manufacturers that evidence standards for cancer products overall are changing, and that confirmatory evidence is not required in some cases. The wider implications of regulatory evidence standards as signal to medicines developers are discussed in section 6.3.

This study suggests that post-marketing obligations imposed by regulators were unlikely to resolve unanswered questions concerning efficacy and safety for a number of reasons.

4.4.4. Issues with timely evidence generation in the post-marketing setting

Firstly, there were often delays, sometimes substantial, in the conduct of studies and submission of results from post-marketing obligations. More than half (61%) of EMA post-marketing obligations and 40% of FDA obligations in the CMA and AA pathways, respectively, were submitted with a delay or were ongoing behind schedule. This is broadly in line with previous studies of EMA post-marketing obligations.[74, 271] For the FDA, a rate of 7% delayed ongoing studies is only half of that found in a 25-year investigation of AA,[155] but this may be explained by the difference in timeframes. Overall, this study adds to previously voiced concerns about the timeliness of post-marketing studies conducted as part of special approval pathways.[61, 70]

4.4.5. Randomised vs. non-randomised study designs in the post-marketing setting

Another reason why post-marketing obligations were unlikely to provide adequate answers to open questions about efficacy and safety is that regulators appear to face a trade-off between receiving confirmatory evidence either from robustly designed

studies (i.e. RCTs) or studies conducted in populations matching the approved indication, but not both. The EMA routinely accepted non-comparative studies to generate confirmatory data, even in cases where CMA was granted on the basis of single-arm trials (bosutinib, brentuximab vedotin, and vismodegib), while FDA post-marketing obligations more typically consisted of RCTs. The lack of a control in non-comparative studies hinders causal interpretation of observed treatment effects. The suitability of such study designs to establish or confirm a positive benefit-risk ratio of conditionally approved medicines is therefore questionable, as is reflected in conventional grading of evidence frameworks and the default study design preferences of regulatory agencies (see section 2.5).[162, 245, 272-274] At the same time, the vast majority (87%) of EMA post-marketing studies were planned to be conducted in the same patient population for which approval was granted, while the FDA accepted confirmatory studies conducted in a different setting for a substantial minority (41%) of cases. This highlights an important difference in the two agencies' approach to post-marketing evidence generation: while the EMA has a demonstrated preference for confirmatory evidence to come from the same population as the approved indication, the FDA allows confirmatory studies to be conducted in a different population (typically in patients with less advanced disease) in an effort to ensure post-marketing obligations are fulfilled.[245] The FDA's policy reflects considerations that RCTs may be difficult or unethical to conduct in rare diseases or populations with no available therapy (although these arguments can be challenged by examples of RCTs being conducted in rare diseases and under challenging conditions[157, 160]). Importantly, approval itself can be turned into an argument against conducting a RCT to confirm the clinical benefit of a new medicine. Regulatory approval signals a positive benefit-risk balance to patients and physicians, even if based on preliminary data, and creates ethical challenges for withholding the treatment from patients when conducting a confirmatory trial in the approved indication in a setting where the treatment has become available. However, failure to conduct a robust study in the intended patient populations results in the availability of a product on the market without confirmed positive benefit-risk ratio; in other words, a different evidence standard being applied where preliminary trial results are deemed sufficient

4.4.6. Surrogate endpoints vs. clinical outcomes in the post-marketing setting

Finally, for the products included in this study, post-marketing studies may also not adequately address uncertainty around the benefit-risk ratio due to the use of surrogate measures, such as response rate or PFS, instead of patient-relevant outcomes, such as overall survival or quality of life. As discussed in section 2.3.2, surrogate measures are potentially a useful tool to shorten clinical trial duration if they reliably predict clinical benefit, yet the validity of even widely used surrogates such as PFS in cancer often remains unproven.[145, 275] When surrogate measures are used for regulatory approval of cancer medicines, they are typically non-validated in both Europe and the US.[144, 152]

In this study, the most commonly used primary endpoint in clinical post-marketing studies of medicines approved through EMA CMA or FDA AA pathways was PFS, followed by response rate. For medicines with less complete efficacy data than usually required at the time of approval, it is therefore unlikely that robust overall survival data become available in the post-marketing setting. Post-marketing obligations included overall survival as primary endpoint for only 25% and 24% of EMA and FDA approvals, respectively. Moreover, all EMA obligations after CMA with overall survival primary endpoints were single-arm studies, which are not suitable for measuring time-to-event endpoints due to the lack of a comparator.

This study adds to other research that shows what type of clinical evidence can be expected to materialise in the post-marketing setting. In the US, confirmatory studies attached to AA overall have been found to mostly measure response rates or PFS, rather than overall survival.[61] Among cancer medicines approvals, the same surrogate measures intending to predict clinical benefit in pivotal trials were used to confirm it in 20% of cases.[155] In Europe, the EMA reported that one third of confirmatory studies for CMAs measured the same type of effect as the pivotal studies (the vast majority being surrogates), with only few studies (12%) measuring overall survival in the post-marketing setting.[156] Overall, this accumulating evidence suggests that, even in the post-marketing setting, the benefit-risk ratio of medicines benefitting from EMA CMA or FDA AA continues to be based on surrogate measures, rather than final, patient-relevant outcomes.

4.4.7. Policy implications

There is now a body of literature, including this study, demonstrating that post-marketing studies are often delayed in both EU and US,[61, 70, 74] do not commonly use robust study designs or patient-relevant endpoints,[33, 61, 85, 155] and are subject to substantial deviations from initially imposed requirements.[70, 271] This study therefore adds to concerns about placing too much emphasis on post-marketing studies to address uncertainties in the evidence submitted at the time of market approval.[74, 276] Given the importance of post-marketing evidence generation as an integral component of EMA CMA and FDA AA, shortcomings in the design and conduct of studies in this setting raise the question whether these expedited approval pathways work as intended.

Evidence standards for conditional vs. regular approvals

This study complements and extends research by others that have highlighted the "exception" in drug approval becoming the norm. Kesselheim et al. found that two thirds of all newly approved medicines in the US are now approved through a special regulatory pathway.[55] This development may have impacted on the evidence threshold for regulatory decisions more widely. In this study of cancer medicines, evidence from the same pivotal trials was submitted to EMA and FDA for 19 of 21 matched indications. While this evidence was only considered sufficient for approval through the CMA or AA pathways by one agency (with a view to collecting additional evidence in the post-marketing setting), it was commonly deemed appropriate for regular approval by the other agency. In many cases, the exception of approving a product with limited evidence on safety and efficacy has therefore become the norm. Policy makers need to be aware of de facto (as opposed to statutory) evidence standards applied by regulatory agencies to enter an informed discussion about whether this is a desired trajectory for regulating the medicinal products market.

The availability of evidence is shaped by how regulators interpret the requirements for market approval of new medicines. Striving not to be perceived as barriers to pharmaceutical innovation and patient access to promising new treatments, regulatory bodies now aim to position themselves as enablers of medicinal product development, as evidenced in strategic plans by the EMA and FDA.[57, 58] For example, when CMA was first introduced in the EU in 2006, regulators used the ambiguity surrounding the

role of this new regulatory tool to turn it into a "rescue" option for market approval applications that would otherwise be rejected.[179] Similarly, granting regular approval to medicines with limited efficacy and safety data can affect the regulatory landscape overall by signalling to pharmaceutical companies that evidence that was historically considered for approval through conditional regulatory pathways (EMA CMA or FDA AA) may be sufficient for regular approval.

The distinction between regular approval and approval through EMA CMA or FDA AA pathways is not trivial: as this study has shown, evidence thresholds set by the regulators through different approval routes translate directly into what evidence becomes available. The products included in this study had uncertain benefit-risk profiles at the time of approval and additional evidence on efficacy and safety typically only became available when a post-marketing obligation was imposed by the regulatory body as part of CMA or AA. Overall, for the 12 indications in this study sample with regular approval by one agency, additional evidence on efficacy was only requested in two cases by the agency granting regular approval. In all other cases, confirmatory evidence that reduces the uncertainty only materialised due to additional studies imposed by the other agency as a condition to granting approval. This suggests that, unless regulators impose the conduct of additional studies (typically through CMA at the EMA or AA at the FDA), the evidence on efficacy and safety in the intended population that is available at the time of market approval is all we will get. It is therefore important for policy makers to consider the potential consequences of a shift in evidence standards for initial approval. The routine granting of full approval for medicines with uncertain benefits and risks can create a precedent for a lower evidence threshold for regulatory approval being applied indiscriminately, with relevant additional evidence unlikely to materialise once the product is on the market.

Addressing issues with post-marketing evidence generation

To address issues with post-marketing study designs, it is instructive to revisit the original AA model, introduced in the US in 1992 to address the evidence vs. access conundrum when promising treatments emerged to treat HIV/AIDS patients. The FDA has described the "ideal approach" to AA as following up pivotal, randomised trials in the post-marketing phase to obtain confirmatory evidence using patient-relevant outcomes after initial approval based on a surrogate measure.[245] However, this

standard was abandoned by the FDA due to concerns about the feasibility of continuing a RCT while the experimental treatment was already available, and a policy to accept studies in similar patient populations as confirmatory evidence was implemented. This policy appears to be effective in incentivising the generation of confirmatory evidence (in the present study, post-marketing evidence generation obligations for 87% of indications approved through the AA pathway had been considered fulfilled by the FDA), but it comes at the cost of not obtaining confirmatory evidence for the approved indication. In these circumstances, it is important to consider the uncertainty surrounding the evidence base for initial (conditional) approval. It is the need for confirmation of an indicative positive benefit-risk ratio that leads to using conditional – rather than regular – approval. Strong confirmatory evidence can be expected to be obtained from robust studies demonstrating a causal positive treatment effect in the population that the medicine is intended for.

A second key issue to be addressed is the timely availability of confirmatory evidence. Policy makers should aim to create a regulatory environment where pharmaceutical companies are incentivised to produce this evidence. Currently, there appears to be little to lose for companies once their product has been approved. While some legal instruments exist in both the EU and the US to enforce compliance with post-marketing obligations (see Table 4), these are rarely used. The threat of revoking approval remains a tame one in the US, where FDA officials appear unwilling to withdraw market approval due to lack of proven efficacy in a lengthy and resource intensive process that invokes an image of a regulator blocking patients from accessing an effective treatment.[65] Stricter enforcement of compliance with post-marketing obligations could improve the rate of timely fulfilled obligations and lead to complete evidence packages becoming available for medicines with EMA CMA or FDA AA. Regulators need to have the mandate to ensure compliance with post-marketing obligations as part of their mission to protect public health and ensure the availability of high quality, safe and effective medicines for citizens.[1, 2] In addition, companies could be incentivised to ensure timely completion of post-marketing studies by making up-to-date status reports on post-marketing obligations and their results publicly and easily available. While such a database has been introduced in the US, it does not appear to be up-to-date and complete (see study limitations below). An equivalent

database for the EU is missing. The status of post-marketing studies and their results are important information for patients, their physicians, and the public. However, this information is often difficult to obtain, leaving concerned patients, physicians and the public in the dark about when complete evidence of the benefits and risks will be available and what it will contain.

4.4.8. Limitations

This study covers cancer medicines with EMA decisions from 2009 to 2013 inclusive. The timeframe was chosen in order to allow for sufficient time to complete post-marketing obligations and take the current status (as of 2018, when this study was conducted) of EMA CMAs and FDA AAs into consideration. The time period allowed for a minimum of five years to complete and submit post-marketing studies, in line with previous research.[33]

The focus on cancer medicines is due to this being by far the largest group of therapeutic agents receiving EMA CMA and FDA AA. In the US, approximately 60% of medicines with AA from 2000 to 2013 were cancer products,[62] and in the EU, over half of medicines with CMA from 2006 to 2010 belonged to that category.[156] Despite the focus on one therapeutic area, the findings are therefore highly relevant for other products that benefit from conditional approval pathways. An example from outside cancer is hydroxyprogesterone caproate, a medicine for preventing pre-term births which had received FDA AA but failed to demonstrate effectiveness in a confirmatory trial, leading consumer rights advocates to call for its withdrawal from the market.[277]

The focus on cancer medicines with regulatory outcomes from 2009 to 2013 led to a sample of 21 indications. Small sample sizes bring about limitations regarding the generalisability of results, but this disadvantage is more than offset by the opportunity to investigate the evidence requirements and regulatory decisions for each included indication in more detail. Focusing on a 5-year period allowed an in-depth investigation of regulatory decision making under uncertainty, including both pre-approval and post-approval evidence requirements. Furthermore, this study included all cancer indications in the chosen time period for which either EMA CMA or FDA AA was granted. The study sample therefore represents a full overview of regulatory decisions for cancer medicines for which at least one regulator considered the evidence on

efficacy and safety insufficient for regular approval. Due to negative FDA decisions not being published systematically, cases where the FDA denied approval and an application for EMA approval was never submitted could not be considered. However, this is in line with the aims of this study, which focused on indications for which a regulatory outcome by both EMA and FDA existed.

There were also limitations in the extracted information. Publicly available documents formed the primary source for data used in this study, but these sometimes lacked important details. For example, for most approvals for variations of previously authorised products, the FDA does not provide review documents. While information on pivotal trials and post-marketing obligations could still be extracted from other publicly available documents (label and approval letter), more detailed information on the decision-making process for these approvals, such as the reasoning given for granting conditional vs. regular approval, could not be extracted.

Finally, there were challenges when ascertaining the status of post-marketing obligations. The FDA maintains a database of post-marketing commitments and requirements. However, this database appears not to be comprehensive and often missed requirements that were listed in FDA approval letters of products included in this study. Information provided in supplemental approval letters was deemed more reliable. The status of post-marketing obligations was therefore not always determined through official FDA documents. The EMA has taken a step towards improving transparency of post-marketing obligations for products with CMA through the release of a report in 2016 that contained details on all obligations.[156] However, this was only a snapshot in time; determining the status of obligations after the cut-off date of that report was challenging. The status of European obligations can be determined by examining updated versions of marketing authorisation documents (specifically, Annexes II), but information on the timeliness of submissions and the results of post-marketing studies are not readily available. Instead, discussions of post-marketing obligations in the EMA's committees, for which meeting minutes are published (although these lack detail), were used as sources and the removal of obligations from the Annexes II was used to indicate that obligations had been considered fulfilled. Due to the difficulties described above for obtaining detailed information about post-marketing obligations for both EMA and FDA, it was only possible to determine

whether obligations were considered fulfilled by the regulatory agency – this included instances where the manufacturer was released of an obligation, and the study was never completed, and at least one case where a trial demonstrated inferiority of the conditionally approved drug,[251] but the obligation to conduct the study was considered fulfilled.

4.4.9. Conclusions

US and European regulatory bodies were often willing to grant regular approval to cancer medicines with uncertain benefit-risk profiles, rather than deny approval or require the collection of additional evidence through AA and CMA pathways, respectively. Non-randomised studies played an important role in both the initial approval of these products and in the post-marketing phase, as the majority of pivotal trials with identified uncertainties were single-arm trials across both EMA and FDA, and non-randomised studies featured in three quarters of post-marketing obligations imposed by EMA in one third of FDA post-marketing obligations. When post-marketing studies were imposed, both their design (i.e. non-randomised studies and / or continued reliance on surrogate endpoints) and their delayed submission of results raise questions over FDA's AA and EMA's CMA ability to reconcile early market access with maintaining rigorous regulatory standards.

5. Regulatory acceptance of non-randomised evidence and emergence of post-authorisation RCTs for cancer medicines in Europe

Chapter summary

The EMA ordinarily requires clinical evidence for approval of new medicines from RCTs, yet marketing authorisations for new anti-cancer medicines are increasingly based on non-randomised studies, specifically single-arm trials (SATs). This chapter analysed the reasons for accepting SATs as pivotal evidence for marketing authorisations of 55 anti-cancer medicines in Europe from 2014 to 2023 and assessed their validity.

While approval based on SAT evidence presents a deviation from the methodological standard for regulatory approval, clear justifications for regulatory acceptance of SAT were often lacking in publicly available documents. Potential reasons for lack of RCT feasibility or necessity were not found to be commonly applicable to the 55 indications: less than 15% reached different criteria for "dramatic" effect sizes which may obviate the need for a RCT, sample sizes required for a two-arm RCT were smaller than the number of participants in pivotal SATs for the majority of indications, and matching RCTs for the approved indications did not take longer to complete than pivotal SATs. Finally, while a RCT was identified for the majority of indications, confirmatory evidence on a statistically significant survival on overall survival only became available for some indications, while the majority of RCTs continued to measure surrogate endpoints.

The findings presented in this chapter indicate that RCTs appear to be feasible for a substantial proportion of cancer medicines approved based on SATs. With RCTs commonly emerging for cancer drugs initially authorised based on evidence from SATs, explicit discussion of the trade-offs between deviation from the gold standard and early authorisation through acceptance of typically more uncertain SAT evidence is warranted.

5.1. Introduction

5.1.1. Regulatory evidence standards

Market approval for new medicines has historically been granted on the basis of robust evidence on the treatment's efficacy and safety, in the form of two RCTs. For the EU, relevant legislation states that "[i]n general, clinical trials shall be done as 'controlled clinical trials' and if possible, randomized; any other design shall be justified."[278] However, the drive to accelerate access to new medicines and to rapidly provide patients with unmet medical need with treatments has led to an increase in the use of other forms of evidence as the basis for regulatory decisions. Expedited approval pathways and designations, such as conditional marketing authorisation and PRIME in the EU and accelerated approval and Fast Track and Breakthrough Therapy designations in the US, allow the submission of preliminary evidence, including from non-randomised studies that do not have a concurrent control group (i.e. single-arm trials, SATs). As described in the literature review (sections 2.4.1 and 2.4.2), SAT results can be submitted on their own or contextualised through comparison with an external control group. This study does not distinguish between these specific cases, as it focuses on regulatory acceptance of non-randomised studies, defined as studies lacking random allocation of participants to treatment and control groups to identify the treatment effect of a new medicine. This is in line with empirical studies which have documented a clear upwards trend in the proportion of regulatory approvals based on non-randomised studies, in particular for cancer medicines (see sections 2.4.1 and 2.4.2 in the literature review).[28, 29, 36, 157, 159]

Compared to double-blind RCTs, SATs are at increased risk of bias.[23] Non-randomised studies have been shown to be more likely to produce misleading results (i.e. results that were overturned by subsequent studies),[94] and to overestimate treatment effects when compared to RCTs.[95, 96, 279] Deviating from the gold standard when seeking marketing authorisation for a new medicinal product should therefore be well justified. For example, RCTs are considered challenging and lengthy to conduct in therapeutic areas with very small patient numbers. The assumption that RCTs are impossible to conduct for rare diseases, however, has been challenged by evidence showing that approximately one third of regulatory trials for new drugs to

treat rare cancers were randomised, and that the likelihood of a regulatory trial being randomised (vs. non-randomised) was independent of annual incidence of the cancer.[160]

5.1.2. Regulatory acceptance of non-randomised studies in prior research

There are potentially valid reasons why RCTs are not feasible or not required for authorisation of new medicines for a given indication, and approval based on data from SATs (including explicit or implicit comparison from external controls, such as historical cohort data) is justified. Approval solely based on SATs may therefore reflect an appropriate flexibility of the regulatory agencies in areas where conducting a RCT would not be feasible or ethical and insisting on it would create a barrier to patients accessing the new treatment.

Given the increase in the use of SATs for regulatory approval, it is important to understand whether deviating from RCTs is justified. While studies have characterised some aspects of regulatory approvals based on SATs in the US and Europe, such as trial characteristics and therapeutic areas,[28, 29, 31, 82, 133, 157, 280] effect sizes,[29, 136] and risk of bias,[158] there are no empirical studies comprehensively mapping the reasons provided for submitting or accepting SAT as pivotal evidence. There is also a lack of studies comprehensively assessing whether RCTs emerge for clinical questions about the benefit-risk profile of medicines that were approved on the basis of SATs only.

5.1.3. Study aims

This study aimed to analyse the reasons provided for accepting SATs as pivotal evidence for marketing authorisations of new anti-cancer medicines in Europe, to assess their validity, and to analyse how often RCTs are conducted after initial marketing authorisation based on SATs.

5.2. Methods

Potential reasons for accepting SATs instead of RCTs for regulatory approval of new medicines were reviewed. For a sample of cancer medicine indications authorised

by the EMA on the basis of SAT evidence from 2014 to 2023, relevant information on the pivotal SAT and the regulatory review was extracted. Potential comparators were identified to calculate relative treatment effects and sample sizes of hypothetical RCTs. Subsequently, actual RCTs were identified that were conducted after the pivotal SATs and characteristics of indications with subsequent RCT vs. without subsequent RCT were compared.

5.2.1. Framework of potential reasons for accepting single-arm studies

Empirical studies and regulatory guidance documents from the EU and the US were reviewed to develop a framework of potentially valid and distinct reasons for submitting (as applicant) and accepting (as regulatory body) non-randomised studies as pivotal evidence for marketing authorisations of new medicinal products (Table 9).

Ethical arguments and technical reasons for not conducting a RCT

The underlying ethical argument for deviating from RCTs for marketing authorisation is to enable or accelerate access to new treatments for patients with unmet medical needs. This ethical argument is typically supported by two types of technical arguments for why non-RCT data on the efficacy and safety of a new treatment should be accepted. First, there are arguments relating to the feasibility of conducting RCTs or the feasibility of completing them in a timely fashion. These concern the size of the patient population and the time it takes to complete a RCT. Second, there are possible objective criteria for when a non-RCT may produce robust evidence that obviates the need for an RCT. These relate to the type of disease (a well-established natural course of disease without available treatments) and a potential large effect estimate observed in the non-RCT. More details on the types of arguments are provided in Table 9.

Table 9: Framework of reasons for deviating from RCTs for regulatory approval

Main arguments	Reasons for not conducting a RCT
RCT is not feasible	*Small patient population:* Low numbers of patients have been described as a situation where "a randomized trial may simply not be possible", e.g. when a therapy is being tested in a subset of patients carrying a specific biomarker.[281] The FDA also states that rarity of disease may

	provide a reason for not conducting a trial with concurrent controls, citing challenges with generating evidence compared to common diseases.[18]
	Time required to conduct a RCT: Single-arm trials may take less time to complete than RCTs for two reasons. Firstly, without a concurrent control group, fewer patients may need to be recruited overall and the target sample size may be achieved quicker (depending on the required sample size as determined through power calculations). This may be particularly relevant when patient numbers are so small that it would take a very long time to recruit enough patients for treatment and control arm and it would be virtually impossible to complete a trial within a time that is acceptable for the development of a therapy. Secondly, single-arm trials are not qualified to measure time-related endpoints that require comparative data from a concurrent control arm, such as overall survival. While this is a limitation which introduces considerable uncertainty about the clinical meaningfulness of any therapeutic effect observed in the trial, the endpoints used in single-arm trials (e.g. tumour response) can be measured quicker than survival, resulting in an earlier trial completion date.
RCT is not necessary	*Well-established natural course of disease without available treatment:* If the natural course of the disease is well-known, a trial without concurrent control may provide convincing evidence of the treatment's effect. Such a situation may exist for severely debilitating or life-threatening diseases without available treatment options (i.e. unmet medical need), where an outcome is measured that is biologically plausible to respond to the investigational treatment and highly unlikely to show spontaneous improvement (e.g. tumour shrinkage). Both FDA and EMA refer to the need for a clearly established natural course of disease, where the outcome of the disease without treatment is predictable, as a pre-requisite for making causal inference about treatment effects from studies without a concurrent control.[16, 18, 23] It should be noted that, in this line of arguing, lack of equipoise about the benefits of the experimental treatment compared to the standard of care is assumed. This results in an ethical argument for not withholding the experimental treatment from some trial participants (e.g. the control group in a placebo-controlled RCT) who have a severely debilitating or deadly disease without available acceptable treatment options (particularly if the new

	treatment has shown promising treatment effects in early phase trials).[31]
	Dramatic effect size: A sufficiently large effect size observed in a non-randomised study could indicate a true treatment effect beyond the impact of bias inherent to such study designs.[282] Rate ratios of 10 or larger were suggested as representing true treatment effects that can be inferred from non-randomised studies,[283] although regulators have not specified a threshold for effect sizes that would render a RCT unnecessary.[133] Internationally harmonised regulatory guidance mentions dramatic treatment effect as potential justification for lack of a concurrent randomised control group.[16]

The reasons outlined in Table 9 provide a technical justification for the ethical argument that different evidence standards should be applied to areas with high unmet medical need, so that patients who have no other (acceptable) treatment option and have a severely debilitating or deadly disease can get access to potentially effective new treatments quicker. Pharmaceutical companies and regulators see SATs as effective means of accelerating the development and approval process for new therapies, in particular when these are conducted in narrowly defined, often small patient populations and demonstrate large response rates that can be attributed to the therapy under investigation.[284, 285] The US FDA explicitly states that it might be appropriate to accept a higher risk of false positive treatment effects (i.e. finding an effect when there really is none) in some situations if this allows potentially effective treatments to become available or become available earlier for patients.[18]

Regulatory guidance

The FDA provides explicit guidance on situations when deviations from the traditional evidence standards may be acceptable, including for diseases with unmet medical need with a highly predictive and severe natural course (leading to mortality or severe morbidity), when the treatment effect is self-evident (e.g. for general anaesthetics), and for rare diseases [18]. The EMA guidance – adopted from ICH – is less detailed, but states that external controls (i.e. lack of a concurrent randomised control) should be restricted to situations when there is a dramatic treatment effect, and where the natural course of the disease is well-known (i.e. spontaneous improvements can be ruled out). Importantly, the guidance refers to ethical issues around conducting

comparative trials for conditions with unmet medical need, but states that even in these circumstances randomised trials are generally preferable when the natural course of the disease cannot be reliably predicted.[16] This is also reflected in EMA guidance on conducting trials in small patient populations.[286] In 2024, the EMA published more detailed reflections on the use of SATs to establish efficacy.[23] This included a working definition of when SATs can isolate treatment effects, i.e. when a causal interpretation of evidence from a SAT is appropriate: "If in a SAT individual outcomes for the planned endpoint are observed that could not occur without effective treatment within the designated follow-up period for any trial participant, the SAT is able to isolate the treatment effect on that specific endpoint."[23, p5] However, the EMA reflection paper also states that this requires qualitative reasoning and the isolation of a treatment effect is therefore always subject to uncertainty. A treatment effect can also only be isolated for an adequate endpoint (i.e. one that does not occur without effective treatment), and the EMA considers that a SAT can only be considered as pivotal evidence when such an endpoint is used. Key challenges for using SATs to establish efficacy also include the confidence in treatment effect estimates in comparison to the counterfactual (which can be impacted by the characteristics of patients included in the trial, without design feature to control for characteristics, as would be the case in a RCT), the external validity (which is also a challenge for RCTs, but exacerbated in SATs if the patient population is heterogeneous but the trial population is not representative), and quantification of uncertainty (variability in outcomes for individual patients which cannot be observed for the counterfactual).

The ICH guidance highlights the potential lack of comparability between study groups as the key issue in trials without a randomised concurrent control. They also include considerations of different types of bias that could be addressed through appropriate trial design, including considerations for choice of control group (including external controls).[16, 192]

Other reasons, such as the cost of running a RCT and the lack of external validity (due to restrictive inclusion/exclusion criteria in trials), should not be considered valid arguments for deviating from the gold standard evaluation for regulatory purposes. These reasons do not feature in regulatory guidance documents on appropriateness of different clinical trial designs for marketing authorisation.[18, 192]

5.2.2. Sample identification

This study included cancer medicine indications for which the pivotal evidence came from a SAT and for which the EMA deemed the benefit-risk profile to be favourable between January 2014 and December 2023. The sample was identified from the EMA's medicines data table, which comprises approved medicines, withdrawn applications and status of opinions. European Public Assessment Reports (EPARs) for all entries for human cancer medicines (ATC codes L01-L04) that were not generic, biosimilar, or hybrid medicines, or otherwise used known active substances, and with a first marketing authorisation date from 1 January 2014 to 31 December 2023 were reviewed. Products were included when they were first authorised in that period and when the pivotal trial was not a parallel-group RCT, i.e. when the pivotal trial (or, in the case of several pivotal trials, none of them) was not a RCT that allocates participants to the new treatment or a comparator such as placebo, active comparator (another medicine), or standard of care.[157] This could include SATs with or without explicit external controls, as well as trials with randomised allocation of participants, but where all of the trial arms receive the investigational treatment. Throughout this study, these trials are collectively referred to as SATs, meaning trials without a concurrent control group that does not receive the new treatment.[14, 85, 157]

5.2.3. Data extraction

Data extraction for characteristics of pivotal trials and subsequent RCTs, as well as potential comparators, outcomes, and prevalence was conducted by one researcher, with verification by a second researcher. A data extraction template and guidance document for each data field were used to ensure consistency. Data extraction for justifications for accepting single-arm trials was conducted by a single researcher.

From EPARs, information on the authorised indication, regulatory pathway (regular, conditional, and marketing authorisation under exceptional circumstances), orphan and PRIME designation, authorisation date, reasons provided for accepting SAT evidence, potential alternative treatments for the indication, and post-marketing study obligations for clinical efficacy and/or safety studies was extracted. For the latter, annexes II of the summary of product characteristics (SmPC) were also reviewed.

For clinical trials (both pivotal SATs and subsequent RCTs), patient eligibility criteria, intervention and (for RCTs) comparator details, endpoints, sample size, trial start and primary (actual or expected) completion dates, and treatment effects for the primary endpoint, and, where available, overall survival, progression- or event-free survival, and overall response rate were extracted. Information on clinical trials was preferably sourced from clinicaltrials.gov; alternative sources (EPAR or publications) were used when information was not available or was unclear from clinicaltrials.gov. Extraction of trial characteristics and results from clinicaltrials.gov meant that there may be discrepancies to the data available for review at the time of initial marketing authorisation (e.g. because the pivotal trial was on-going while results were submitted for marketing authorisation). There may therefore be discrepancies in the number of participants treated and the results. Data from clinicaltrials.gov are preferred because they show the trial as it was completed, representing the most mature evidence in relation to the newly approved product.

Incidence data for the condition for which a product was approved was used as a proxy indicator for the number of patients eligible for treatment. Granular incidence on subgroups of cancers was extracted or calculated where this was relevant for the approved indication (e.g. information was extracted on the proportion of patients with specific gene mutations when this was included in the approved indication), using the 2022 GLOBOCAN annual cancer incidence estimates per 100,000 population for Europe,[287] which were extracted from Cancer Today (https://gco.iarc.fr/today/). Incidence data was available for cancer types (e.g. lung, bladder, thyroid, multiple myeloma, leukaemia). Targeted literature searches were conducted for specific gene mutations or other patient characteristics as specified in the indications of products included in the study sample. Incidence data for each cancer type were then multiplied with the proportion of patients matching the approved indication in the sample to obtain an estimate of the incidence of the relevant patient population. The most recent incidence estimates (i.e. mostly 2022) and any available estimate for proportion of patients with specific characteristics for a given cancer type were used, assuming that incidence was stable.

5.2.4. Identifying relevant comparators

Potential comparators for included cancer medicine indications were primarily identified from the EPARs and technology appraisal guidance documents from the National Institute of Health and Care Excellence (NICE), the English HTA body. Differently from regulatory review, HTA is used to assess the added therapeutic value of a new product compared to the standard of care. Standard of care can vary across markets. Assessments conducted by a single HTA body were chosen for a consistent standard of care, and NICE was selected because information on potentially suitable comparators were available in English language both for published appraisals as well as for products for which only scoping of the appraisal was conducted. Existing treatment alternatives discussed by the European Society for Medical Oncology (ESMO) and in peer-reviewed scientific studies that previously identified potential comparators for products approved based on SATs were also reviewed.[30, 136, 158, 159, 288] For each comparator, effect estimates were extracted for overall or objective response rate (ORR) and, where available, overall survival (OS). Treatment options may be limited for multiple refractory indications, and there may not be a clear standard of care. Therefore, potential comparators where only included when they were discussed as relevant treatment options by both EMA and NICE (52.7% of included indications), or – in cases where NICE did not publish an assessment, did not provide a clear comparator, or identified a different comparator than EMA – by EMA and ESMO or published literature (7.3%), or when clearly considered by EMA as a relevant comparator (14.5%). When no clear option could be identified from the sources listed above, no comparator was included (25.5%).

5.2.5. Identifying emerging RCTs

RCTs for the included indications were identified through systematic searches in clinicaltrials.gov (using the expert search function to search for the name of the pharmaceutical in combination with randomised assignment in the trial and the condition) and – when no RCT was identified in clinicaltrials.gov – in MEDLINE via PubMed (searching for the name of the pharmaceutical in combination with a highly sensitive filter for clinical trials). RCTs were eligible for inclusion if they studied the same medicine for broadly the same therapeutic indication (e.g. lung cancer). For

sensitivity analyses, the match between identified RCTs and pivotal SATs was assessed based on the compatibility of patient eligibility criteria: pivotal SAT and subsequent RCT were deemed compatible if patients from the pivotal trial would also be considered eligible for the RCT. This could happen when there were no or only negligible differences in eligibility criteria between pivotal SAT and subsequent RCT or when there were differences, but the RCT eligibility criteria encompassed those of the SAT (e.g. the RCT was conducted in patients with two or more prior treatments, but the pivotal trial only included patients with three or more prior treatments). Compatibility assessments were conducted by a single researcher. Independent assessment for 15% of the sample by a second researcher showed agreement for 87.5% of assessments.

When several potentially matching RCTs were identified, the included RCT was selected based on whether it was completed or had results (primary criterion) and based on the compatibility of patient eligibility criteria.

5.2.6. Analysis

The primary unit of observation were medicine indications. The study reported the number and proportion of indications where pre-defined and other reasons for accepting SATs were mentioned in EPARs.

For each indication with identified comparator, the effect size for a hypothetical comparative trial was estimated, following the approach by Djulbegovic et al.[136] Extracted data on the interventional treatment and control were used to estimate odds ratios (OR) and risk ratios (RR). 2x2 tables were created with the number of patients with a response (as defined in the primary endpoint of the pivotal SAT) and total number of patients in the SAT, and the number of patients with a response and total number of patients for the comparator. Comparator data could come from SATs or the individual arms of comparative trials. When more than one potential source trial existed, a conservative approach was applied to extract the most beneficial effect estimate. When no clear comparator could be identified, no effect size was estimated and the indication was excluded from analyses relying on relative effect estimates.

Statistical power calculations were conducted on the basis of observed effect size for each SAT and effect size of identified comparators to obtain the sample size required to demonstrate the observed treatment effect with 80% statistical power and a

two-sided α of 0.05. Hypothetical sample size for a two-sided RCT was calculated using the following formula:[289]

$$n = \frac{(Z_{1-\alpha/2} + Z_{1-\beta/2})^2 * [p_1(1-p_1) + p_2(1-p_2)]}{(p_1 - p_2)^2}$$

P_1 and p_2 are the proportions of responding patients for the novel intervention and the comparator, respectively, and the Z-values are the critical values of the normal distribution for α (confidence level) and β (statistical power). The sample size of the hypothetical RCT was compared to the sample size of the pivotal SAT for each indication, and the median difference across indications was then calculated. When there were several pivotal trials per indication, only the largest trial was used for this analysis. In using the observed effect size from a SAT to calculate the sample size required for a comparative trial, no adjustments were made to factor in potential differences in expected treatment effects and populations, resulting in a potential underestimation of the required sample size.[290, 291]

The effect estimates derived from pivotal SATs were compared to three thresholds that have been proposed in the literature to determine whether they constituted "dramatic" treatment effects, following the approach by Djulbegovic et al.[136] for analysing large treatment effects: OR ≥ 12 (i.e. an empirically derived threshold from a review of marketing authorisations based on non-randomised evidence),[136] RR ≥ 10 (i.e. a threshold above which effect sizes are considered unlikely to reflect anything other than actual treatment effects),[283] and RR ≥ 5 (i.e. the threshold above which the GRADE framework for grading quality of evidence considers study design to be very unlikely to explain associations).[131] Sensitivity of this analysis to the calculated effect estimates was assessed by increasing and decreasing these by 50%.

For each indication, the duration (in months) of pivotal trials and RCTs was calculated (using the median when there was more than one pivotal trial per indication). Trial duration was then summarised separately for pivotal trials and RCTs using the median and interquartile range (IQR) and the difference in median duration was calculated. The Wilcoxon signed-rank test was used to compare the distribution of trial duration between pivotal SATs and RCTs.

The study tested whether the proportion of products with post-authorisation RCTs was statistically significantly different from those without post-authorisation RCTs for a

range of characteristics including reasons for accepting SAT for marketing authorisation. Chi-squared tests were used to compare proportions when all expected cell counts were ≥5, and Fisher's exact test when any expected count was <5.

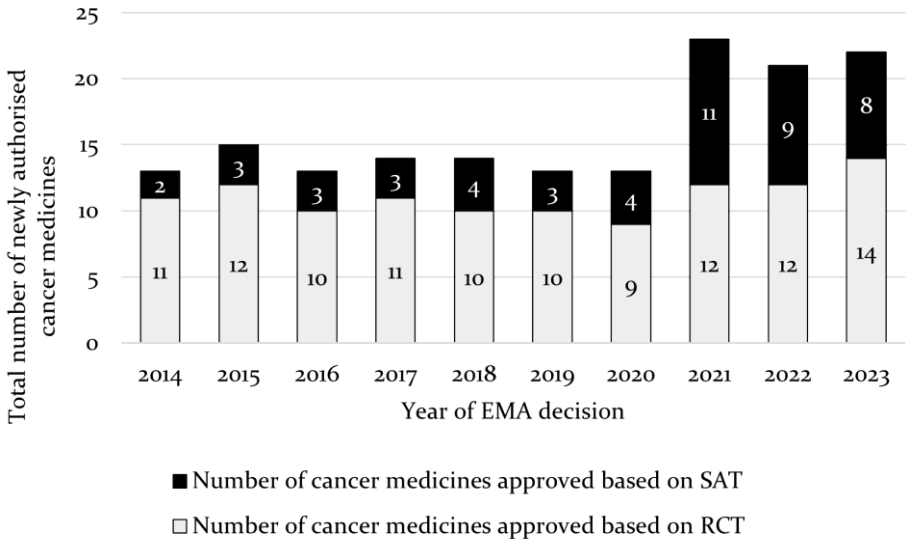
Finally, effect estimates of actual RCTs were analysed by categorising their results for overall survival, progression- or event-free survival, and overall response rate according to whether a statistically significant benefit relative to the comparator was demonstrated.

5.3. Results

From 2014 to 2023, the EMA authorised 161 new cancer medicines, including 50 (31.1% of all new approvals) for which the pivotal evidence on benefits and risks came from a SAT (Figure 13). The proportion of new cancer medicines approved based on SAT data increased from below one quarter between 2014-2017 to between one third to almost half of all new cancer medicines between 2021-2023.

Of the 50 included products, one was withdrawn from the market after authorisation (moxetumomab pasudotox for treatment of hairy cell leukaemia). For four products, the EMA approved distinct indications based on different pivotal trials or different clinical considerations and these were included as separate indications in this study. The sample in this study therefore consisted of 55 cancer medicine indications (see Appendix 5 for details of included indications).

Figure 13: Number of new cancer medicines per year and pivotal trial design



Abbreviations: RCT, randomised controlled trial; SAT, single-arm trial.

5.3.1. Regulatory pathways and pivotal trial characteristics

Characteristics of included indications are shown in Table 10. Among this sample of products authorised exclusively based on pivotal SATs, 47 indications (85.5%) were assessed based on a single pivotal trial each, five indications (9.1%) were assessed based

on two pivotal trials, and three indications (5.5%) were assessed based on three pivotal trials. 61 unique trials were used as pivotal trials. Of these, one trial provided two unique trial arms and another trial provided three unique trial arms, for a total of 64 unique single-arm trial cohorts. The median number of enrolled patients in pivotal SATs or in the relevant arm of a multi-arm non-comparative trial was 111 (IQR, 79.75 to 149.5). All pivotal trials measured some form of response rate as primary endpoint. The most common endpoints were overall response rate (35 trial arms, 54.7%) and objective response rate (25 trial arms, 39.1%). Some of the pivotal trials included co-primary endpoints, such as duration of response or individual components of overall response, such as complete response or intracranial overall response. For these trials, only the use of overall or objective response was reported.

Table 10: Characteristics of cancer medicine indications authorised by the EMA on the basis of SAT evidence from 2014-2023

Characteristic	**n**	**%**
Total number of cancer medicine indications	55	100
Cancer type		
Haematological oncology	29	52.7
Solid tumour oncology	26	47.3
Marketing status		
Authorised	54	98.2
Withdrawn post-approval	1	1.8
Regulatory approval pathway		
Regular	17	30.9
Approval under exceptional circumstances	2	3.6
Conditional marketing authorisation	36	65.5
Orphan designation		
Yes	24	43.6
No	31	56.4
PRIME designation		
Yes	12	21.8
No	43	78.2
Post-authorisation study obligations included a RCT		
Yes	36	65.5
No	19	34.5
Number of pivotal trials per indication		
1	47	85.5
2	5	9.1
3	3	5.5

5.3.2. Reasons for accepting single-arm trials

There was notable variation in how the reasons provided for accepting evidence from SAT as pivotal trial, rather than RCT evidence, were discussed in the EPARs. Exemplary quotes for the reasons included in the *a priori* developed framework, as well as for other reasons, are provided in Table 11. Dramatic effect size in some form was most commonly referenced (n=29, 52.7%), followed by an assessment that the natural course of the disease was well established and that there was no available treatment (n=24, 43.6%; Figure 14). The other two main reasons for accepting non-RCT evidence from the *a priori* developed framework were only rarely mentioned (small patient population: n=8, 14.5%; time required to conduct RCT: n=4, 7.3%). For 36 indications (65.5%), two or more reasons were identified.

Reasons for accepting SAT evidence were sometimes explicitly stated and in other cases had to be inferred from text in the EPARs relating to the SAT design. For example, effect sizes were not consistently labelled as "dramatic", but reference to the observed treatment effect featured in the majority of EPARs, e.g. by being described as "outstanding" (EPARs for dostarlimab, ibrutinib, larotrectinib, rucaparib, among others), "compelling" (EPARs for cemiplimab, idecabtagene vicleucel and rucaparib, among others) or "substantial" compared to alternatives (EPAR for mosunetuzumab), by discussing this as demonstrating clinical benefit or being likely to result in long-term clinical benefit (EPARs for belantamab mafodotin, loncastuximab tesirine, sotorasib, tepotinib, among others), or by referring to effects observed for other treatments (EPARs for ciltacabtagene autoleucel, lisocabtagene maraleucel, and osimertinib).

Figure 14: Reasons for accepting single-arm trials as pivotal evidence

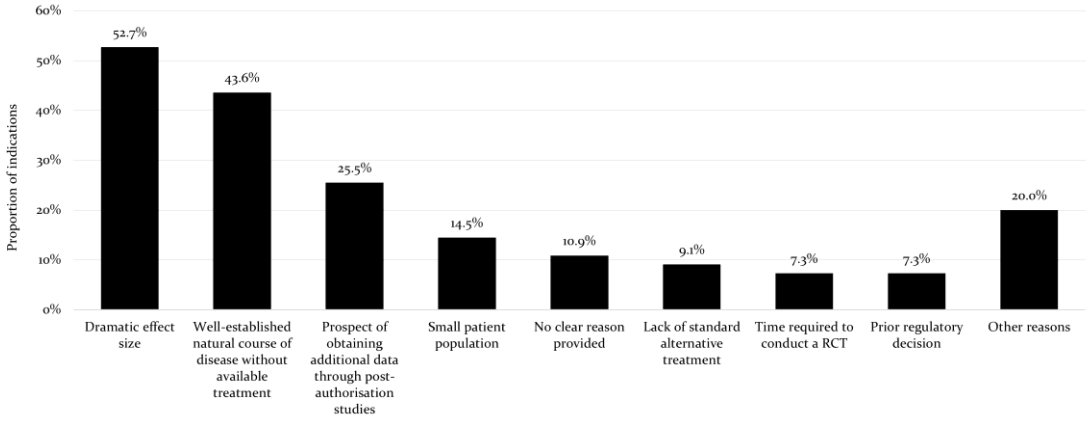


Table 11: Exemplary quotes for reasons for accepting single-arms trials provided in EPARs

Reason	Exemplary quotes	Indication
Small patient population	"In view of the recent experience, a randomized study is considered a challenge in this disease setting. This is mainly due to the fact that the size of the population available for clinical studies is limited and there is no single standard treatment option among the several dozen available ones that are used according to different investigator practices and individualised treatment approaches, which in turn would require large sample sizes in order to ensure sufficient power to detect important treatment effects."	Idelalisib (non-Hodkin lymphoma)
	"The CHMP agreed that the applicant is unable to provide comprehensive data on the efficacy and safety under normal conditions of use because the indication is too rare. Therefore, the CHMP recommended that a marketing authorisation under "exceptional circumstances" be granted due to the rarity of the disease."	Moxetumomab pasudotox (hairy cell leukaemia)
	The small patient population was discussed by members of the Scientific Advisory Group (SAG), and its reasoning was then taken up by the CHMP: "Although there was general agreement that a clinical benefit has been established some members regretted that lack of a randomized controlled trial did not allow a robust comparison to available intensive chemotherapy regimens. However, the majority of SAG members considered that RCTs, although theoretically possible, would have been extremely challenging due to the rarity of the disease and acknowledged that study 0114 is the only prospective trial with pre-specified response criteria performed in this rare disease. Some SAG members noted that in the post marketing setting, more than 120 patients have been treated with tagraxofusp within a rather short time suggesting that a small randomized trial vs. physician's choice would have been possible despite the rarity of this disease. Others, however, noted that many centres would have to be opened to achieve even a small sample	Tagraxofusp (blastic plasmacytoid dendritic cell neoplasm (BPDCN))

	size and that this is not realistic given that most centres will not open a clinical trial when they expect to include only 1-2 patients/year." "The CHMP agreed that the applicant is unable to provide comprehensive data on the efficacy and safety under normal conditions of use because the indication is too rare. Therefore, the CHMP recommended that a marketing authorisation under "exceptional circumstances" be granted due to the rarity of the disease."	
Time required to conduct a RCT	Given that a RCT in the first-line setting was under way at the time of the marketing authorisation application (CMA), the applicant used a time element in its justification for applying for a conditional marketing authorisation (supporting the immediate availability of the product): "Postponing the submission of the MAA until completion of the ARROW study would lead to a delay of approximately two years in time to approval. A delay of two years is excessive in the setting of a fatal disease with an important medical need, considering metastatic NSCLC patients in a 1L setting have a median OS of approximately 10-30 months (Keytruda SmPC, 2021; Taxotere SmPC, 2020), and patients in a 2L treatment setting have a median survival time of approximately 8-13 months (Alimta SmPC, 2020; Fehrenbacher et al, 2018; Herbst et al, 2016; Opdivo SmPC, 2020)." While the EMA did not explicitly refer to this argument in its justification for granting a CMA, it appears to have followed the reasoning given that it considered the benefits to public health of immediate availability to outweigh the risks: "The benefits to public health of the immediate availability outweigh the risks inherent in the fact that additional data are still required. Given the positive benefit/risk and the unmet medical need in the applied indications as described above, this is considered fulfilled."	Pralsetinib (non-small cell lung cancer)
	As part of the requirements for granting a CMA, the EMA has to consider whether immediate availabiltiy of the treatment outweighs the risks to to public health given the uncertainty. For this product, the applicant explicitly mentioned the time component in relation to collecting data in a RCT. While the EMA did not comment on this issue specifically, it did accept that public health benefits of immediate availability outweigh the risks. "Although the pivotal	Teclistamab (multiple myeloma)

	study adopted a Phase 1/2 design that did not utilise randomisation, the applicant considers that a delay to gather further or comparative data would be disproportionate from a public health perspective, as teclistamab addresses an unmet medical need and offers a new therapeutic option for heavily pretreated patients with multiple myeloma, which outweighs the risks due to the immediate need for further data."	
	"Data on post-BCRi therapy are limited, therefore confirmation of efficacy is needed. This patient population is expected to increase as more CLL patients will (eventually) progress after BCRi treatment. A comparative clinical study in this setting is likely not to be feasible within an acceptable time frame considering that the BCRis have only been rather recently approved relative to the long PFS/DoR of these products as this is a requirement for a conditional approval. Sixty more patients are planned for enrolment in study M14-032; the extension of this supportive study will form the basis of additional efficacy data needed."	Venetoclax (Chronic lymphocytic leukemia (CLL))
Well-established natural course of disease without available treatment	"A single-arm trial design can be appropriate for a setting where there is no approved or acceptable therapeutic option. In the 1L cisplatin ineligible patients, there is no established standard of care, and having in mind the poor prognosis in these patient, it may be justified not randomizing patients to "physicians best choice"."	Atezolizumab (urothelial carcinoma)
	"The pivotal study is a single arm study with no comparator. However, as ALK is a driver mutation in ALK-positive NSCLC and lorlatinib is a targeted treatment against a well-known target, a randomisation to chemotherapy, immunotherapy or placebo is not considered ethical in a situation, when the disease progresses after use of the available ALK-inhibitors, due to the poor efficacy of these treatment options. Furthermore, all of the developed ALK-inhibitors have been superior in efficacy compared to chemotherapy and immunotherapy so far, both regarding ORR and duration of response and there is an unmet medical need in patients who progressed after second generation ALKi in the 2nd or 3rd line setting."	Lorlatinib (non-small cell lung cancer)
	"The primary efficacy endpoint ORR supported with an IRC-based sensitivity analysis is acceptable in the context of an uncontrolled trial. […] Although the single arm design could be acceptable for the TCR population considering the initiation date of the study and absence	Melphalan flufenamide (multiple myeloma)

	of a standard of care at that time, recent approval of alternative treatment options emphasizes the need for contextualization of the efficacy data, in particular for patients that are TCR after relative few prior lines of therapy. Moreover, uncertainty remains with regard to the effect on time-dependent endpoints, PFS and OS, which cannot be reliably interpreted in an uncontrolled study."	
	"The primary endpoint [ORR] is acceptable in SAT for regulatory purpose where there is no alternative therapy with proven benefit in terms of PFS or OS, since this is a clear reflection of the pharmacodynamic effect (antitumour activity)."	Selinexor (multiple myeloma)
	"The single-arm, open-label design of the pivotal study U201 is acceptable in the context of a CMA, since the targeted indication is for the last-line setting and there is no SOC." "As the basis for this application is a single-arm trial, several aspects need explicit consideration. In this respect, the natural course of the target disease as well as the target population are considered adequately defined."	Trastuzumab deruxtecan (breast cancer)
Dramatic effect size	"Avapritinib has shown an outstanding and durable ORR in patients expressing the PDGFRA D842V mutation regardless of prior line of therapy, which is unprecedented in a population subset that traditionally is unresponsive to TKI therapy. Even acknowledging the limitations associated to the low number of patients treated, the clinical benefit of this treatment in the intended target population has been shown."	Avapritinib (gastrointestinal stromal tumors)
	"Concerns were expressed by the CHMP (EMEA/H/SA/3117/7/2019/PA/ADT/PR/III) that the ORR cut-off was not an a priori decision and that an ORR above 25% would not be considered sufficiently compelling in the context of a single pivotal, single arm trial. These issues, however, are considered overcome by the high response rates reported for KTE-X19 (ORR 85%, 95% CI 75.0%, 92.3%), indicating significant results would be achieved also with a substantially higher ORR success criterion."	Brexucabtagene autoleucel (mantle cell lymphoma (MCL))
	"Overall, the main limitations of the study are the non-randomised nature and the absence of control arm leading to uncertainties on the actual benefit in clinically relevant treatment	Dostarlimab (endometrial cancer)

	outcomes, i.e. OS. Furthermore, preliminary efficacy remains limited to 108 patients with dMMR/MSI-H advanced EC. These limitations could be acceptable in the context of a CMA based on the observed effect on ORR when it is considered outstanding and provided all the criteria for the CMA are fulfilled (see further discussion under efficacy data and additional analyses)."	
	"The main limitation of this clinical trial is its non-randomised nature and the lack of a control arm, in addition to the fact that this is the only clinical trial in support of this submission. These limitations could be considered acceptable by the CHMP in the context of a CMA when the effect observed is considered outstanding and all the requirements for a CMA are fulfilled."	Selpercatinib (non-small cell lung cancer and thyroid cancer)
	"This response rate is considered promising and much higher than expected beforehand (~35%) based on literature data. Sensitivity analyses support the reported ORR. The magnitude of response is considered highly clinically relevant for this heavily pre-treated study population, who have no standard of care. It is noted that very few had a non-evaluable best overall response, which is reassuring. The waterfall plot clearly reflects these data, which are considered highly clinically relevant for this heavily pre-treated study population, who have no standard of care. Hence, the ORR data show clinically significant activity of trastuzumab deruxtecan even though no comparative data is available." "The substantial longer DoR together with the high ORR support a major therapeutic advantage despite the inherent limitations related to cross-study comparisons. The difference is as such that it is considered to overcome the remaining uncertainties related to indirect comparisons and high censoring rate."	Trastuzumab deruxtecan (breast cancer)
Prospect of obtaining additional data through post-	"Nevertheless, uncertainties on the actual clinical benefit remain due to the immaturity of data and lack of comparator. Furthermore, the safety database is small and limited in terms of long-term exposure. These uncertainties will be addressed by means of the specific obligations which will provide comprehensive efficacy and safety data. In conclusion, it is considered that preliminary efficacy and safety data support a favourable benefit/risk balance	Dostarlimab (endometrial cancer)

authorisation studies	for dostarlimab in the treatment of patients with advanced dMMR/MSI-H EC. Nevertheless, additional comprehensive evidence will be provided in order to establish firm conclusions on the actual benefits and risks of dostarlimab (see SOBs)."	
	"The study was designed as a Phase 1 dose-finding and dose escalation study without the required rigor usually requested for confirmatory trials (adequate pre-specification of key elements, appropriate type 1 error control, blinded and independent decision making, randomized control arm, etc) and was amended 10 times changing its nature from a Phase 1 to a Phase 1/2 (and finally "confirmatory") study. As comprehensive data on the product are not available, a conditional marketing authorisation is agreed. In the context of the specific obligations the applicant will provide the updated study report for study NP30179. The applicant will submit the results from a phase III study GO41944 as a specific obligation."	Glofitamab (B-cell lymphoma)
	"Efficacy has been established on the basis of durable ORR in a single-arm trial. Although the durable response is considered a clinically meaningful benefit, there is a need to further quantify the efficacy of pemigatinib in a comparative trial. Study INCB 54828-302, a phase III study to evaluate the efficacy and safety of pemigatinib compared with gemcitabine plus cisplatin in the first-line treatment of participants with FGFR2-rearranged cholangiocarcinoma is ongoing."	Pemigatinib (cholangiocarcinoma)
Lack of standard alternative treatment	"The phase 2 study was conducted without an active control arm which is a limitation. Because there is no standard treatment for these patients, the missing control arm may be considered acceptable."	Belantamab mafodotin (multiple myeloma)
	"JCAR017 was only investigated in single-arm trials (SATs). In principle, randomised trials powered to detect superiority in terms of time-to-event endpoints are required to assess clinical benefit in non-Hodgkin lymphomas (NHLs); however, when the clinical development of JCAR017 was planned, no proper active control could be identified in the target population. Moreover, aggressive lymphomas are rapidly progressive neoplasms, and evidence on the encouraging anti-CD19 CARTs activity in NHLs was already available at the time, so it can be	Lisocabtagene maraleucel (B-cell lymphoma; follicular lymphoma; mediastinal neoplasms)

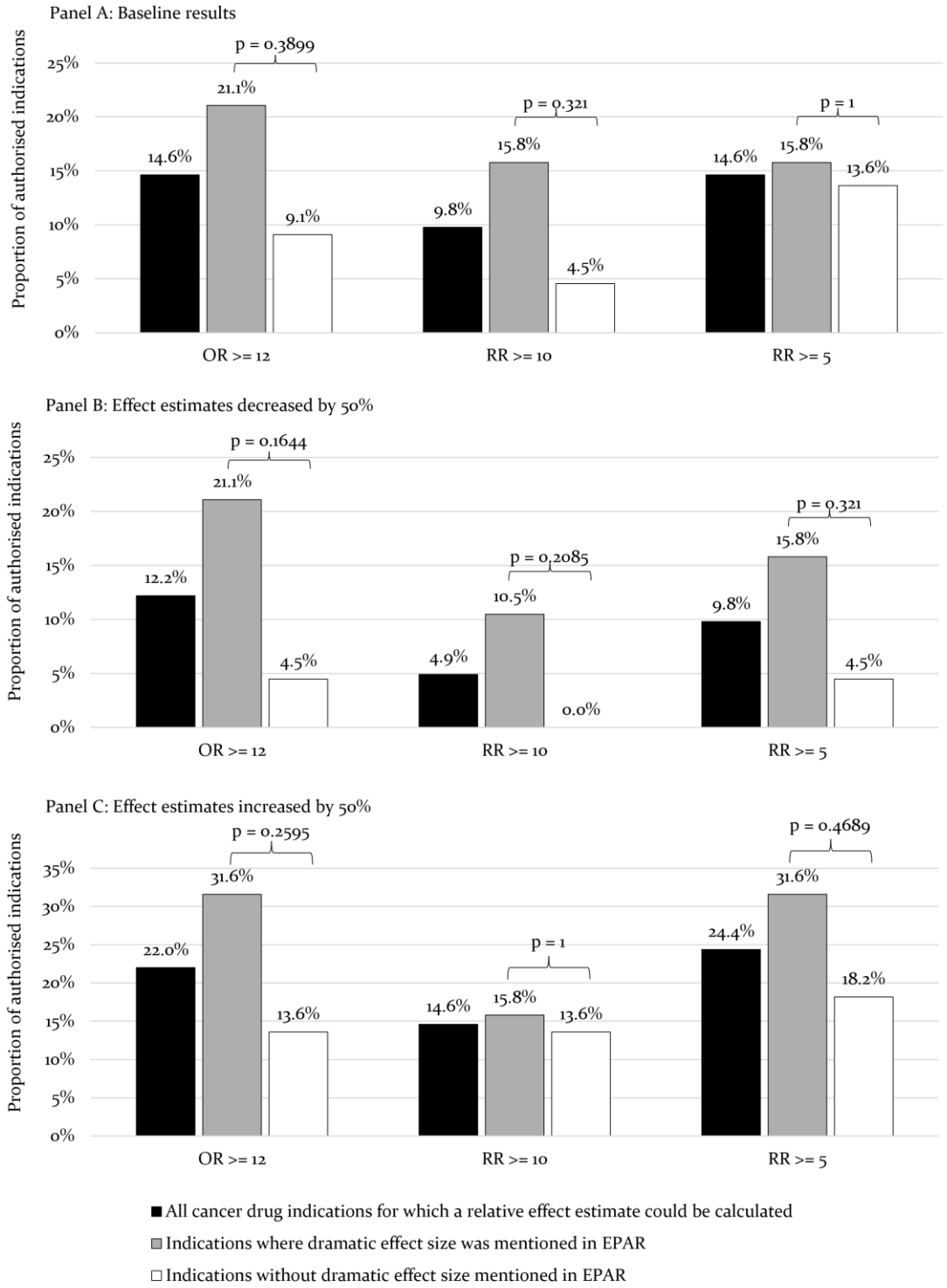
	recognised that placebo or cross-over designs could have been challenging from an ethical perspective.”	
	"One limitation of the study is the single-arm design, since the phase 1/2 study was conducted without an active control arm. Because there is no standard treatment for these patients, the missing control arm may therefore be considered acceptable."	Teclistamab (multiple myeloma)
Prior regulatory decision	“This application is based on a single arm trial which has been accepted for the purpose of MAA in scientific advice, and similar evidence base has supported several recent conditional marketing authorisations. Nevertheless, the evidence for efficacy generated in a single arm trial is less robust and subject to different types of bias, most notably selection bias. Time-to-event endpoints are considered important for demonstration of clinical benefit, but cannot be reliably assessed in a SAT setting. Thus, the pivotal SAT can be considered adequate to demonstrate clinical benefit in this patient population within the context of a CMA, but not to provide comprehensive data.”	Elranatamab (multiple myeloma)
	“This SOB is a replication of the single-arm efficacy (and safety) data in a new and independent single-arm study cohort in order to provide a comprehensive overall data package. For clarity, the term ‘single-arm’ also encompasses studies that contain more than one arm, but do not randomise to a control for a formal comparison, i.e. non-randomised trials as well as trials in which only experimental arms are randomised, but without formal comparisons between the arms (text taken from draft reflection paper single-arm trials). With the term ‘randomized, controlled trial’ (‘RCT’) is meant here a study that does include a formal comparison to a (standard-of-care) control. Such independent replication of single-arm data by a separate single-arm cohort has previously been accepted to provide comprehensive data resulting in a full approval.”	Futibatinib (cholangiocarcinoma)

	"Study MMY1001 was an early, uncontrolled, exploratory trial; in principle, B/R evaluations for Marketing Authorisation Application (MAA) procedures would require high-quality confirmatory data from at least one controlled, randomised trial. However, conditional approvals in advanced settings of MM based on promising results from single-arm trials (SAT) are not unprecedented."	Talquetamab (multiple myeloma)
Other reasons	COVID-19 pandemic: "The applicant did initiate a phase III RCT in November 2020, and is planned to enrol 90 patients according a 2:1 randomisation. This trial in the second-line setting compares capmatinib versus docetaxel (NCT04427072 - study A2301).[...] The difficulty for performing a RCT in the 2L setting is acknowledged given [...], the COVID pandemic [...]"	Capmatinib (non-small cell lung cancer)
	Data available from a RCT in a slightly different population: "Upon consultation, the SAG-O concluded that melflufen + low dose dex is associated with clinically relevant efficacy, with the exception of the subgroup of patients with relapse within 36 months following high-dose melphalan and autologous SCT. In addition, the SAG-O considered that although the exact effect size cannot be determined due to differences in disease and treatment characteristics, the results of study OP-103 obtained in patients of whom most had fewer lines of treatment than the OP-106 patients, are relevant for the target population in study OP-106."	Melphalan flufenamide (multiple myeloma)
	Supply shortage of established treatment alternative: "A comparative trial versus the authorised Erwinia derived asparaginase crisantaspase, the only alternative treatment option, would have been preferred to establish efficacy of recombinant crisantaspase produced in Pseudomonas fluorescens (RC-P). However, due to the repeated recent supply shortages experienced with crisantaspase and well-established intermediate endpoint for efficacy (asparaginase activity), the single arm pivotal trial design was accepted by CHMP."	Asparaginase erwinia chrysanthemi (recombinant) (acute lymphoblastic leukemia and lymphoblastic lymphoma)
	Totality of evidence: "Reassurance around the methodological limitations was provided by the consistency of the results from various additional analyses and across different subgroup populations. Further confirmation of these results is expected with the ongoing randomised study 20019 [BRUIN-MCL-321]."	Pirtobrutinib (mantle cell lymphoma (MCL))

Relative effect estimates could be calculated for 41 of the 55 indications (74.5%), for 19 of the 29 indications that referenced dramatic effect size as justification for accepting SATs (65.5%), and for 22 of the 26 indications that did not reference dramatic effect size in the EPAR (84.6%). Appendix 6 provides details on identified comparators. Among indications where a relative treatment effect could be calculated, the proportion fulfilling three different criteria for dramatic effect sizes was between 9.8% (for RR ≥ 10) and 14.6% (for OR ≥ 12 and for RR ≥ 5) (Figure 15, Panel A). It was numerically higher (ranging from 15.8% to 21.1% for the different thresholds) among indications where dramatic effect size was mentioned in the EPAR compared to when dramatic effect size was not mentioned in the EPAR (4.5% to 13.6%), but the difference in proportions was not statistically significant at conventional levels for any of the three thresholds. In a sensitivity analysis, the proportion fulfilling the three criteria increased slightly compared to the baseline analysis (to 14.6% for RR ≥ 10, 24.4% for RR ≥ 5, and 22.0% for OR ≥ 12) when increasing the calculated effect estimates by 50% (Figure 15, Panel C).

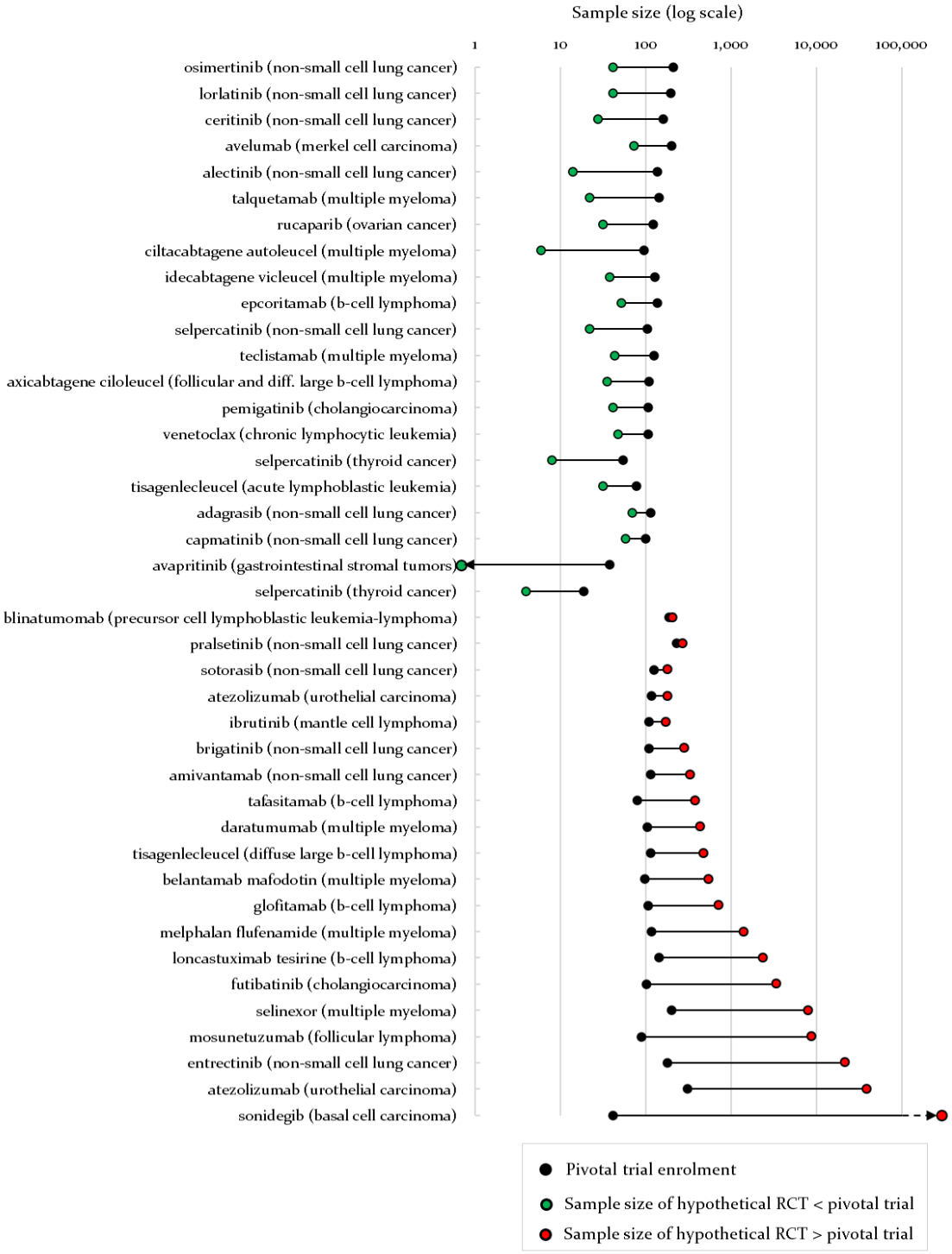
Using the effect estimates of identified comparators, required sample sizes for hypothetical RCTs of the novel anti-cancer products were calculated (Figure 16). The required total sample size for a two-arm RCT (i.e. including both intervention and control groups) was smaller than the number of participants treated with the new product in the pivotal trial for 21 indications (51.2% of indications for which a comparator could be identified). The median sample size of pivotal trials included in this analysis was 115.5 (IQR, 97.5 to 158.5) and median sample size for hypothetical RCTs was 74 (IQR, 36 to 472). The median difference in sample size between pivotal SAT and hypothetical RCT was 15 (IQR, -357 to 83).

Figure 15: Proportion of cancer medicine indications with dramatic effect sizes



Legend: Figure shows the proportion of cancer medicine indications with dramatic effect sizes according to three different thresholds (OR ≥ 12, RR ≥ 10, and RR ≥ 5). Panel A shows results for all included indications for which a relative treatment effect could be calculated (n=41). Panels B and C show sensitivity analyses. Abbreviations: EPAR, European Public Assessment Report; OR, odds ratio; RR, risk ratio.

Figure 16: Sample sizes of pivotal trials and hypothetical RCTs

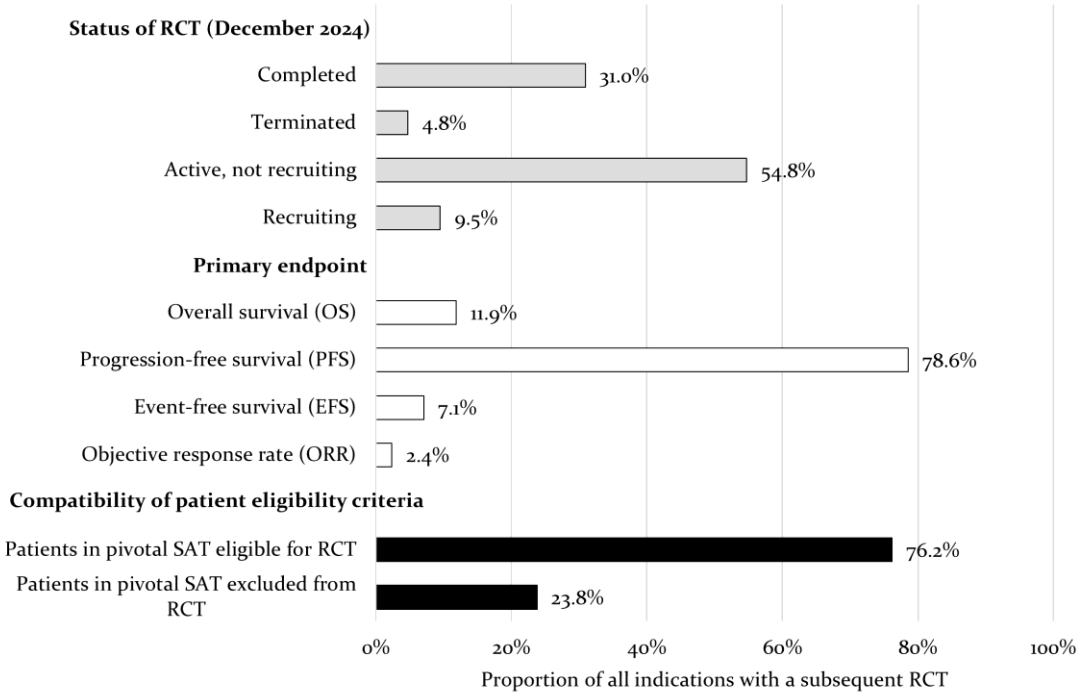


Legend: Figure shows the number of patients enrolled in the pivotal trial for included cancer medicine indications (black circle) and the number of patients required to conduct a hypothetical two-arm RCT for the same indication (green and red circles). The black lines show the absolute difference between actual sample sizes of pivotal single-arm trials and calculated sample sizes of hypothetical RCTs.

5.3.3. Emerging RCTs

A matching RCT (i.e. a RCT of the same product for broadly the same therapeutic indication) was identified for 42 of 55 (76.4%) included cancer medicine indication and no RCT was identified for 13 (23.6%). Among 42 indications with identified RCT, 13 (31.0%) had a completed RCT by December 2024, but the majority of RCTs was ongoing (Figure 17). For 36 of 42 (85.7%) indications with identified RCT, this was required as a post-marketing obligation by the EMA. For most, the subsequent RCT used progression-free survival as primary endpoint (n=33, 78.6%), and for only five indications (11.9%), a RCT with overall survival as primary endpoint was identified. In the majority (76.2%) of identified RCTs, patient eligibility criteria would have allowed patients enrolled in the pivotal SAT to be included.

Figure 17: Characteristics of RCTs identified for indications authorised based on SATs



Abbreviations: RCT, randomised controlled trial; SAT, single-arm trial.

Characteristics of indications were generally not statistically significantly associated with the emergence of a RCT after approval, including cancer type, PRIME designation, reasons provided for accepting a SAT for initial marketing authorisation, as

well as rarity of disease. RCTs were identified at a similar rate for indications with (n=20, 47.6% of indications with identified RCT) and without orphan designation (n=22, 52.4%), and for ultra-rare (incidence <1 per 100,000, n=8, 19.0%), rare (incidence 1-6 per 100,000, n=18, 42.9%), and non-rare indications (n=16, 38.1%) (Table 12). The only two characteristics that were significantly associated with emergence of a RCT were post-marketing obligations for carrying out a RCT and regulatory pathway. While RCTs were identified both for indications where a RCT was imposed as post-marketing obligation (n=36, 85.7% of all authorised indications with identified RCT) and indications without such obligation (n=6, 14.3%), none of the 13 indications without subsequent RCT had a post-marketing obligation for carrying out a RCT. These results did not change in a sensitivity analysis comprising only RCTs with patient eligibility criteria that comprised eligibility criteria for the pivotal SATs.

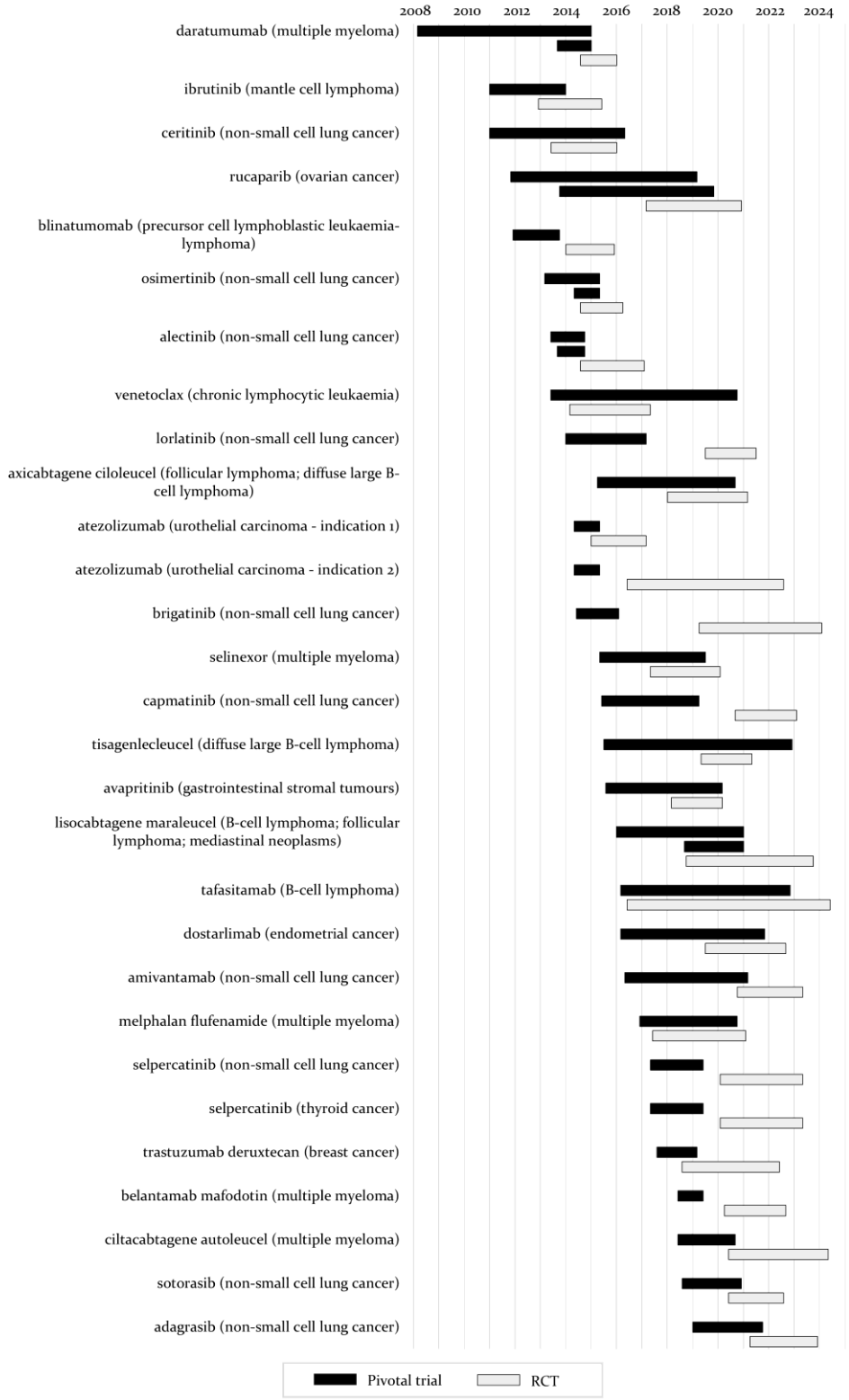
Table 12: Comparison of characteristics between indications with vs. without subsequent RCT

Characteristic	**RCT identified (n, %)**	**No RCT identified (n, %)**	**p-value for comparison between groups**
Indications included in the study	42 (76.4)	13 (23.6)	-
Cancer type			
Haematological oncology	22 (52.4)	7 (53.9)	1 (Chi-Square Test)
Solid tumour oncology	20 (47.6)	6 (46.2)	
Regulatory approval pathway			
Regular	12 (28.6)	5 (38.5)	0.0288 (Fisher's Exact Test)
Conditional marketing authorisation	30 (71.4)	6 (46.2)	
Approval under exceptional circumstances	0 (0.0)	2 (15.4)	
Orphan designation			
Yes	20 (47.6)	4 (30.8)	0.4529 (Chi-Square Test)
No	22 (52.4)	9 (69.2)	
Incidence			
< 1 per 100,000	8 (19.0)	3 (23.1)	0.6088 (Chi-Square Test)
1-6 per 100,000	18 (42.9)	7 (53.8)	
> 6 per 100,000	16 (38.1)	3 (23.1)	
PRIME designation			
Yes	9 (21.4)	3 (23.1)	1 (Chi-Square Test)
No	33 (78.6)	10 (76.9)	

Post-authorisation study obligations included RCT			
Yes	36 (85.7)	0 (0.0)	<0.0001 (Fisher's Exact Test)
No	6 (14.3)	13 (100.0)	
Reason provided for accepting SAT as pivotal evidence			
Dramatic effect size	20 (47.6)	9 (69.2)	0.2955 (Chi-Square Test)
Well-established natural course of disease without available treatment	18 (42.9)	6 (46.2)	1 (Chi-Square Test)
Prospect of obtaining additional data through post-authorisation studies	12 (28.6)	2 (15.4)	0.5555 (Chi-Square Test)
Small patient population	5 (11.9)	3 (23.1)	0.5835 (Chi-Square Test)
No clear reason provided	5 (11.9)	1 (7.7)	1 (Chi-Square Test)
Lack of standard alternative treatment	4 (9.5)	1 (7.7)	1 Chi-Square Test)
Time required to conduct a RCT	4 (9.5)	0 (0.0)	0.5624 (Chi-Square Test)
Prior regulatory decision	3 (7.1)	1 (7.7)	1 (Chi-Square Test)
Other reasons	8 (19.0)	1 (7.7)	0.5905 (Chi-Square Test)

For 29 indications (69.0% of indications with identified RCT), trial duration data (start and primary completion dates) were available. Among these, identified RCTs started after the pivotal trials for all indications, and the RCT started while the pivotal trial was ongoing for 21 (Figure 18). The median trial duration was 38 months (IQR: 36) for pivotal SATs and 32 months (IQR: 18) for RCTs (difference: 6 months; p-value for Wilcoxon signed-rank test of difference in distributions = 0.3143). Duration of the pivotal SAT was longer than the identified RCT for 16 indications (55.2% of indications with trial duration data for SAT and RCT). In a sensitivity analysis including only RCTs with patient eligibility criteria that would also allow participants from the pivotal SAT to be enrolled, the difference in median trial duration was 13 months (median for pivotal trials: 44 months vs. median for RCTs: 13 months; p = 0.1526 for test of difference in distributions).

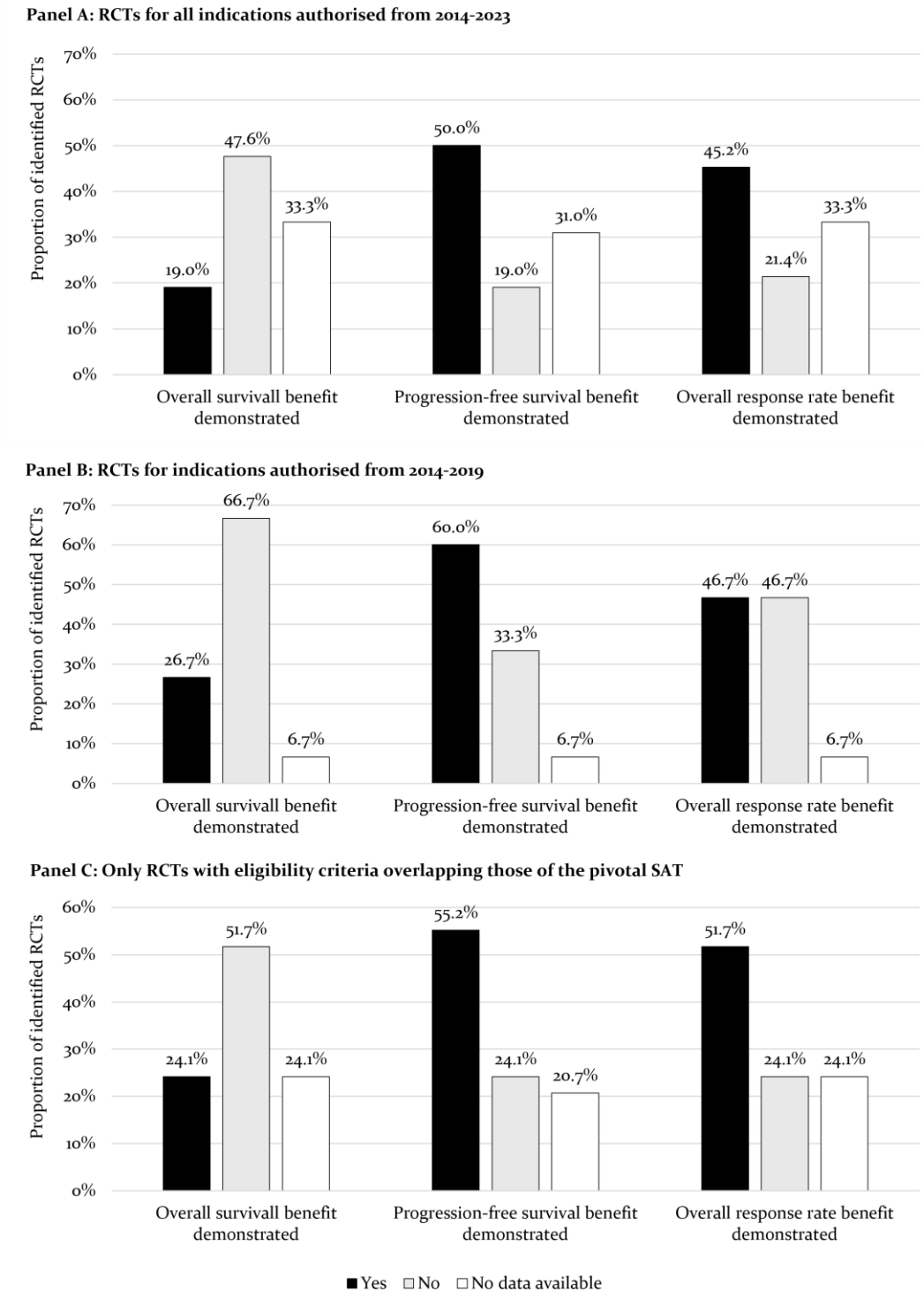
Figure 18: Timelines of pivotal trials and identified RCTs with available start and primary completion dates



Legend: Figure shows start and primary completion dates for pivotal single-arm trials (black bars) and subsequently identified RCTs (grey bars) for the same medicine indication.

By December 2024, overall survival data were available for 28 of 42 identified RCTs (66.7%) and a statistically significant benefit of the novel drug was demonstrated in eight of these (19.0% of identified RCTs, and 14.5% of all indications authorised by the EMA from 2014-2023; Figure 19, panel A). For progression-free survival, data were available for 29 RCTs and a statistically significant benefit was demonstrated for 21 of these (50.0% of identified RCTs; 38.2% of all indications), and for response rate, data were available for 28 RCTs and a statistically significant benefit was demonstrated for 19 of these (45.2% of identified RCTs; 34.5% of all indications). When restricting the sample to indications authorised before 2020, the proportion without available data decreased, but the relative proportion of RCTs demonstrating benefits for overall survival and progression-free survival vs. not demonstrating benefits was similar to the overall sample (Figure 19, panel B). In a sensitivity analysis including only RCTs with patient eligibility criteria that would also allow participants from the pivotal SAT to be enrolled, the proportion of indications with demonstrated benefits did not materially change compared to the baseline (Figure 19, Panel C).

Figure 19: Comparative benefits demonstrated in subsequent RCTs by December 2024



Legend: Figure shows the proportion of indications with RCT evidence demonstrating superiority of the newly authorised treatment for three outcomes (overall survival, progression-free survival, response rate). Panel A shows the proportions for all included indications with identified RCTs, and Panels B and C show sensitivity analyses.
Abbreviations: RCT, randomised controlled trial; SAT, single-arm trial.

5.4. Discussion

5.4.1. Summary of main findings

This study reviewed all new anti-cancer medicines authorised by the EMA on the basis of pivotal evidence from SATs – rather than RCTs – from 2014 to 2023. A key finding was that reasons for accepting SATs as pivotal evidence were not consistently and explicitly stated in public regulatory review documents. While large treatment effects observed in SATs were referenced for more than half of the indications, less than 15% showed "dramatic" effect sizes according to three different thresholds. Finally, the study showed that RCTs appeared to be feasible: for around three in four indications authorised based on SAT evidence, a subsequent RCT was identified. A key predictor for whether a RCT emerged was the inclusion of a post-marketing obligation for conducting a RCT at the time of the initial marketing authorisation.

A strength of this study was the comprehensive identification of potential reasons for accepting SAT evidence, rather than the gold standard RCT. Potential reasons can be grouped into RCTs not being feasible (or not being feasible in a timely fashion) and RCTs not being necessary to obtain credible evidence of a treatment effect. Empirical evidence, including novel insights from this study, raises questions over the validity of these as justifications for deviating from RCTs for marketing authorisation of new anti-cancer medicines.

5.4.2. Feasibility of RCTs

The argument that RCTs are not feasible due to low patient numbers is likely to gain importance in the future. Authors from the EMA have referred to feasibility constraints for conducting RCTs as developers are increasingly seeking approval for products targeting biomarkers.[284] However, small patient numbers may not preclude a RCT in many cases, as shown by the present study and previous research: the number of patients required for a hypothetical two-arm RCT in this study was smaller than the number of patients enrolled in the pivotal SAT for 51.2% of indications. This analysis was conducted ex-post and used observed treatment effects to determine required sample sizes; ex-ante planning of a RCT may lead to larger required sample sizes since this may take potentially optimistic results of feasibility studies and early clinical trials

into account.[290-292] However, these results are in line with previous studies showing that the number of eligible patients does not predict feasibility of RCTs. One study reviewing US FDA approvals of new medicines to treat rare cancers found that approximately one third of regulatory trials were randomised and that the likelihood of a regulatory trial being randomised was independent of annual incidence of the cancer.[160] In Europe, more than half of EMA marketing authorisations for orphan-designated cancer medicines were RCTs.[293] Furthermore, the German HTA body, IQWIG, found that 75% of all orphan medicinal product approvals by the EMA from 2001 to 2013 were based on RCTs, often double-blind.[161]

This study provides novel evidence on how often RCTs emerge for cancer medicines that were initially approved based on pivotal evidence from SATs. A RCT after marketing authorisation based on SAT evidence was identified for approximately three quarters of indications (when applying stricter matching criteria, a RCT was still identified for more than half of the indications). As described in more detail in the literature review section of this thesis (section 2.5.2), previous research focused on RCTs required to confirm clinical benefit following accelerated approval (in the US) or conditional marketing authorisation (in the EU) based on single-arm studies,[30, 230] or discussed emerging RCT evidence after authorisation based on SATs anecdotally.[136, 157] The present study provides novel insights by showing that RCTs are commonly conducted after approval based on SATs in the same or related indications, and that this is not limited to expedited approval pathways, as products with regular marketing authorisation accounted for more than one quarter of identified RCTs. This study also demonstrated the importance of regulatory post-marketing obligations: the vast majority of identified RCTs (including for products with regular authorisation) were imposed by regulators. This is in line with previous research, including a qualitative study that identified incentives for industry to conduct a trial (such as regulatory requirements) as facilitating factor for post-authorisation RCTs.[294]

The median trial duration was six months longer for pivotal SATs compared to subsequent RCTs. While there was no statistically significant difference in the distribution of trial duration, indication-level analysis shows that pivotal SATs took longer to complete than subsequent RCTs for just over half of the indications. Previous

research has not explicitly addressed the question of the time required to conduct SATs vs. RCTs for approvals of oncology indications. For medicines with FDA accelerated approval, median duration times were longer for post-approval studies than for pivotal trials, but the study did not report on clinical trial characteristics (i.e. whether pivotal and post-approval trials were randomised or not).[295] Another study assessed the feasibility of RCTs for solid-tumour products approved based on SATs and estimated the time needed for patient accrual of hypothetical RCTs, concluding that RCTs would have been possible to conduct within comparable timeframes as SATs for at least 80% of solid-tumour medicines authorised by the FDA on the basis of SAT evidence.[159] The present study confirms these findings empirically.

5.4.3. Necessity to conduct a RCT

Given that RCTs may be feasible for the majority of indications authorised based on SATs, it is important to consider whether RCTs are necessary. From a methodological perspective, the limitations of a SAT may be acceptable when important sources for risk of bias are addressed. In its reflection paper on SATs, the EMA lists ten different types of bias and measures to address them.[23] While the EMA reflection paper provides methodological guidance for conducting SATs in a more rigorous manner, it does not explicitly discuss when a RCT is required and when a SAT may be sufficient.

The debate about the necessity of RCTs is not limited to the medicines regulatory setting. Research methodologists have discussed at great length the merits and drawbacks of randomised vs. non-randomised studies, focusing on concerns about the risk of bias inherent to the latter and confidence in their effect estimates. Within this debate, the size of treatment effects plays a role, as "dramatic" effect sizes observed in non-randomised studies have been argued to indicate a true treatment effect, allowing causal inference from non-randomised studies.[283] Indeed, effect sizes obtained from non-randomised studies appear to impact how certain regulators are about treatment effects. Both in the US and the EU, effect sizes obtained from non-randomised studies supporting approval of new medicines were significantly larger for approvals where the regulator did not require a confirmatory RCT, as compared to approvals for which additional RCT data were required.[136, 137]

While larger effect estimates obtained from SATs can reassure regulators at the time of initial marketing authorisation, they may not translate into clinical benefit. Different thresholds have been proposed for what constitutes a "dramatic" effect size providing credible evidence of a treatment effect from a non-randomised study. Applying three different thresholds to this study's sample of cancer medicines authorised based on SAT evidence, they were reached for 15% or less of included indications. These figures are considerably lower than what previous research showed for the same three thresholds for all medicines approved by the EMA based on SATs from 1995 to 2015.[136] Possible explanations include the focus of the current study on cancer medicines, and the quickly evolving treatment landscape in that therapeutic area, with new treatments showing higher response rates being introduced over the study period and acting as comparators, which may result in smaller comparative benefits.

Importantly, the analysis was based on response rates rather than patient-relevant endpoints such as overall survival or quality of life. Unless response rates have been validated as surrogate measures for patient benefit for the specific indication, it is unclear how meaningful the effects measured using these surrogates are for patients, even if they are considered numerically "dramatic". For any given indication, a specific threshold may be considered clinically relevant. However, justifications for these thresholds have been found to not be transparently documented for approximately half of solid cancer products approved by the EMA using SAT evidence.[29]

Finally, whether a RCT is necessary for marketing authorisation also needs to be considered from an ethical perspective. Supporters of early market access for new treatments question whether it is ethical to withhold a potentially effective new treatment from patients who may have no other satisfactory alternative. This argument rests on the assumption that the new treatment is indeed effective. However, at the time of marketing authorisation based on SAT evidence, a treatment effect would typically be shown for a surrogate measure (i.e. response rate). The present study adds to a body of literature showing that, for the majority of cancer medicines first authorised based on SAT, no evidence from RCTs demonstrating improved overall survival emerges. In this study, a statistically significant benefit on overall survival was shown for approximately one fifth of indications for which a RCT was identified. Even

after a follow-up period of at least five years after marketing authorisation, overall survival benefits were only demonstrated for approximately one quarter of indications with subsequent RCT. These findings extend those of previous studies assessing clinical benefit of medicines authorised based on SATs, which were summarised in section 2.5.2. Cohort studies of cancer medicines approved by the FDA on the basis of SATs through the accelerated approval pathway showed that less than one quarter were converted to regular approval based on confirmatory RCTs measuring overall survival.[30, 142] Similarly to the current study, overall survival benefits were only demonstrated for a minority of indications for which a subsequent RCT after approval based on SAT data was identified. For Europe, evidence from RCTs has been used to convert from conditional to regular marketing authorisation for five of 18 cancer medicines authorised based on SATs from 2012 to 2021.[29] That study also showed that only three of 21 analysed SATs provided evidence on "substantial benefit" as per ESMO-MCBS score.[29]

5.4.4. Policy implications

Marketing authorisation decisions have traditionally been based on evidence from two independent RCTs. A deviation from this gold standard would therefore be expected to be well justified. Indeed, the EMA states in a reflection paper on the use of single-arm studies that pharmaceutical companies aiming to base their marketing authorisation application on evidence from such a study are responsible for justifying this deviation from the gold standard.[23] The need for a justification of a deviation from establishing the benefit-risk-ratio based on controlled clinical trials (preferably, randomised) is also stated in EU legislation. However, the evidence presented in this chapter shows that reasons for accepting SATs as pivotal evidence were often not explicitly addressed in the EMA's assessment reports. Given that the EMA (in its guidance) is asking companies to provide justifications for proposing SATs as pivotal trials, these should be transparently documented for the public, and deviations from the legally mandated standard should be explicitly discussed in public regulatory review documents. The EPAR has continuously evolved over time, providing more structured information to transparently document the reasoning behind issuing marketing authorisations (e.g. through the addition of explicit discussion of specific aspects of the benefit-risk balance, including uncertainties in the evidence, and an effects table).

There is scope for continuing this development by explicitly discussing why a SAT was acceptable as pivotal evidence and what measures have been taken to safeguard against risk of bias inherent to this study design.

The motivation for deviating from the evidentiary gold standard for marketing authorisations (i.e. RCTs) is to provide faster access to new treatments for patients in need. Given the increase in marketing authorisation applications based on SAT data,[28, 29, 159] regulators need to strengthen their framework for regulatory acceptance of weaker evidence. This involves explicit consideration of the trade-offs between faster market access and more uncertain evidence. For conditional marketing authorisations, the EMA is required to do so already, as it can only issue a marketing authorisation if the benefits of immediate availability of the product outweigh the risks due to limited data (among other criteria). In principle, weaker pivotal trial evidence could be addressed through rigorous post-marketing studies, as implemented in conditional approval pathways (conditional marketing authorisation in Europe, accelerated approval in the US) which allow regulatory bodies to mandate confirmatory studies and, in principle, to ensure compliance through the threat of potential withdrawal of marketing authorisation. However, research has shown that post-marketing evidence often suffers from the same methodological weaknesses as pivotal trials, that products remain on the market even when their benefit is not confirmed, and that there are delays until post-marketing studies become available.[74, 154, 155, 230, 296, 297] Regulators have recognised some of the challenges associated with conditional approval pathways and are trying to improve adherence to and timely availability of post-marketing obligations, including by requiring confirmatory trials to be underway at the time of granting conditional approval.[297] Notwithstanding these efforts, the evidence standards for initial approval remain essential for ensuring that the evidence required to make informed decisions in clinical practice is available as new treatments enter the market.

Moreover, limitations in pivotal trials inherent to the non-comparative design of SATs are not restricted to conditional approvals. The study presented in this chapter, and others before it, show that SATs also form the basis for regular approval of cancer medicines in both the US and Europe, with less regulatory power to mandate rigorous post-marketing studies.[133, 230, 298] The evidence from SATs used for initial

marketing authorisation may therefore remain the only evidence available for some products (in the present study, no subsequent RCT was identified for one quarter of products authorised based on SAT evidence), or it is the only evidence for some time (i.e. until RCT evidence emerges).

Timely availability of robust evidence on clinical benefit is important for clinical practice and routine availability of new treatments. This requires contextualisation of treatment effects against the standard of care. From a regulatory perspective, data showing clinical benefit over comparators is not necessarily required to demonstrate a favourable benefit-risk profile, but professional societies, HTA bodies and health care payers face challenges in assessing the added value and the role of new products when such data are not available. In the past, withdrawal of products from the market has been used as a measure of the proven benefits and safety of medicines approved based on SAT evidence.[299] In the sample of this study, one of the 50 products authorised by the EMA based on SAT evidence was withdrawn from the market (moxetumomab pasudotox for treatment of hairy cell leukaemia). In the US, five of 72 indications approved based on SAT data from 1999 to 2020 [30] and one of 31 approved from 1973 to 2006 were withdrawn.[299] However, product withdrawal is an extreme measure to safeguard public health. The relevance of new products in clinical practice is determined by their clinical benefit. Information to assess clinical benefit often remains limited, as confirmatory trials for products authorised based on uncertain evidence often do not provide timely and methodologically robust evidence on clinically meaningful endpoints.[155, 230]

5.4.5. Limitations

A limitation of this study is that potential reasons for acceptance of SATs were only identified from published documents. It is possible that specific reasons for accepting or refusing SATs are discussed more explicitly by the EMA committees but that this is not documented comprehensively in the published EPARs. Such nuances may be revealed through qualitative research outside the scope of this study. While internal discussions may not be documented in these reports, the EMA does publish assessment reports for products with withdrawn marketing authorisation applications after unfavourable initial assessment by the EMA. During the study period (2014-2023), there

were three such cases where the pivotal evidence came from a SAT, providing some additional insight into reasoning behind accepting or refusing SAT evidence. In these reports, the appropriateness of a SAT study design was discussed in terms of the effect size not being considered outstanding, a surrogate endpoint not being validated, and feasibility of conducting a RCT given recruitment success for an expansion cohort of a SAT.

Another limitation lies in the hypothetical nature of calculating treatment effects and sample sizes for hypothetical RCTs. This part of the study required identification of a comparator for indications where the standard of care can be dynamic. Comparators were identified from regulatory review documents and assessments made by the English HTA body and may therefore not reflect the standard of care across the EU.

Finally, the likelihood of RCTs emerging for the indications included in this study was potentially underestimated. Firstly, the study only included one RCT (the best match) for each indication. In some cases, several RCTs were identified, but only the best match for the pivotal single-arm trial was included. For example, for belantamab mafadotin for treatment of refractory or relapsed multiple myeloma, at least six RCTs were identified, but only one focusing on monotherapy in the initially approved indication was included, and others investigating various combination therapies including belantamab mafadotin for refractory or relapsed multiple myeloma, or where belantamab mafadotin was used as part of the standard of care, were discarded. This was also the case for other indications, e.g. tafasitamab for relapsed or refractory diffuse large B-cell lymphoma. In other cases, several RCTs were identified, but some showed differences in patient eligibility criteria compared to the pivotal trial (e.g. both lorlatinib and brigatinib for ALK-positive advanced non-small cell lung cancer). Secondly, the study focused on completed RCTs. Additional RCTs that were still active at the time of data collection were also identified (but were not documented systematically). For example, for dostarlimab for treatment of endometrial cancer, two ongoing RCTs with primary completion dates in 2026 and 2027, respectively, were identified (NCT06023862; NCT05201547). While not all of these RCTs would be expected to be completed (e.g. they might be terminated due to slow patient accrual or changing commercial strategies of their sponsors), the number of RCTs for the indications included in this study is therefore likely to increase over time. Thirdly, some

RCTs might have been missed. While systematic searches were conducted in clinicaltrials.gov and – when no RCTs were found there – in MEDLINE via PubMed, there were no systematic searches of registries in other countries.

5.4.6. Conclusions

RCTs emerged for approximately three quarters of cancer medicines authorised based on SAT evidence by the EMA, showing that RCTs are feasible for the majority of cancer indications. The time and sample size required to run RCTs may not be different from SATs. Despite RCT evidence being considered the gold standard for approval of new medicines, deviations through approval based on SATs are commonly not explicitly justified in regulatory documents in Europe. Explicit discussion of the trade-offs between rigorous evidence standards and early authorisation through acceptance of typically more uncertain SAT evidence is warranted.

6. Discussion and policy implications

6.1. Summary of key findings of this thesis

This thesis aimed to investigate the methodological assumptions underpinning increased interest in non-randomised studies for regulatory approval of medicines and how regulators manage uncertainty introduced through non-randomised studies. It makes novel contributions to both methodological and regulatory policy literature by providing empirical assessments of the internal validity of effect estimates obtained from non-randomised studies, regulatory approaches to managing uncertainty across the US and Europe, and potential justifications for regulatory acceptance of single-arm trials.

Key findings for each of the three research questions stated in section 2.6.3 are briefly summarised below.

Research question 1: internal validity of effect estimates from non-randomised studies

How similar are treatment effects for new medicines obtained from non-randomised studies compared to RCTs?

The first research question was addressed in chapter 3, which presents the results of a meta-epidemiological study of 346 clinical questions including 2,746 contributing individual studies. Across all 346 meta-analyses of pharmacological interventions, the study did not find strong evidence that non-randomised studies systematically over- or underestimated the treatment effects obtained from RCTs. However, the lack of a systematic difference in effect estimates at the aggregate level masked important discrepancies regarding treatment effects generated by the two study types. For a substantial proportion of clinical questions, the difference in effect estimates between randomised and non-randomised studies was two-fold or more (35% of meta-analyses) or beyond chance (16%), and the statistical conclusions about the existence of a beneficial or detrimental effect differed depending on which study type provided the evidence (38%). Moreover, there was evidence for a systematic overestimation of beneficial treatment effects by 19% in interventional non-randomised studies compared to RCTs.

This study quantified the uncertainty about effect estimates obtained from non-randomised studies. It showed that relying on evidence from non-randomised studies alone may lead to substantially different conclusions about whether medicines work compared to when evidence from RCTs is available. Caution is therefore warranted when relying on non-randomised studies as substitutes for RCTs.

Research question 2: regulatory management of uncertainty

Which tools are used by regulatory bodies to manage uncertainty in the clinical evidence when assessing new medicines for market approval, and how do they impact on the cumulative evidence available on benefits and harms of new medicines?

The study presented in chapter 4 compared regulatory outcomes and clinical evidence across both pre- and post-marketing settings for 21 cancer medicine indications assessed by both EMA and FDA from 2009 to 2013. Pre-approval evidence packages for all 21 indications were less complete than normally required, as indicated by the granting of conditional – rather than regular – approval by one of the two agencies, and consisted of data from single-arm trials for the majority of indications. The study revealed that regulators apply a range of tools to manage such uncertain evidence, including the decision to grant or refuse approval, grant conditional approval with the statutory power to request confirmatory evidence after approval, request post-marketing studies under both conditional and regular approval, and restrict the wording of the approved indication.

This study highlighted that regulators are afforded some flexibility to manage uncertainty and that they are more likely to use this flexibility to grant – rather than withhold – approval even when the evidence is less complete than what would usually be expected. In fact, for more than half of the included indications, either EMA or FDA granted regular approval, i.e. they forewent the statutory power of conditional approval to withdraw marketing authorisation should the clinical benefit not be confirmed.

The study also showed that post-marketing evidence was more likely to emerge when the initial approval was conditional than regular, highlighting the importance of regulatory pathways for the cumulative evidence on clinical benefit. However, confirmatory studies were often unlikely to fully address remaining uncertainties, as

studies imposed by both EMA and FDA showed methodological weaknesses. FDA more commonly requested confirmatory RCTs but these were typically conducted in earlier treatment settings than the approved indication, while EMA commonly accepted single-arm or observational studies as confirmatory evidence. Non-randomised evidence featured in post-marketing obligations for three quarters of EMA approvals and in one third of FDA approvals. Across both EMA's and FDA's post-marketing obligations, confirmatory evidence on overall survival was only expected for a small minority of indications, and confirmatory studies were often delayed.

While conditional approval therefore appears to impact on the availability of post-marketing evidence, the methodological quality and timeliness of that evidence raises questions over how well conditional approval pathways are able to reconcile robust evidence standards with early access to new medicines.

Research question 3: regulatory acceptance of non-randomised studies

What are the reasons for regulators to accept non-randomised studies for marketing authorisation of new medicines, and are these reasons justified?

The study presented in chapter 5 reviewed 55 cancer medicine indications approved by the EMA from 2014 to 2023 for which the pivotal evidence came from single-arm trials. Despite single-arm trials representing a clear deviation from the traditional standard for medicines approval, the study showed that publicly available regulatory review documents often did not explicitly discuss the reasons for accepting these study designs rather than RCTs.

The study assessed potential reasons for accepting single-arm trials but found that these were unlikely to explain regulatory acceptance for the 55 approved indications. The first of these reasons – a large treatment effect in a non-randomised study that may obviate the need for a RCT – was found to be true for less than 15% of indications. Another – small patient populations rendering RCTs unfeasible – was unlikely to be relevant since the sample sizes of pivotal single-arm trials were often larger than the sample sizes that would be required for a hypothetical two-arm RCT. Indeed, actual RCTs were identified for approximately three quarters of the 55 indications, and there was on average no significant difference in how long it took to complete these trials,

showing that RCTs were often feasible to conduct in a similar time frame as single-arm trials. While RCTs did emerge, they commonly did not provide confirmatory evidence on overall survival, and even confirmation of benefit based on surrogate endpoints was only achieved for approximately half of the indications.

The study thus showed that RCTs are feasible for the majority of cancer indications that were first authorised based on single-arm trial data. The time and sample size required to run RCTs may not be different from single-arm trials. When conducted, RCTs commonly failed to demonstrate overall survival benefit for these cancer medicines, raising questions over patient benefit of early access based on phase II trial data.

Taken together, the empirical findings of this thesis indicate that increased reliance on non-randomised studies for regulatory approval leads to important challenges in understanding the clinical benefit of new medicines for patients. While expedited approval pathways – in particular conditional approval – aim to reconcile early access to medicines with rigorous evidence standards, the uncertainty introduced by non-randomised studies is unlikely to be resolved over the life cycle of new medicines.

6.2. Benefits and risks of using evidence from non-randomised studies for medicines approval

As described in the introduction of this thesis (section 1.4), regulators aim to balance the benefits of early approval for new medicines with maintaining rigorous evidence standards. Authorising medicines on the basis of non-randomised studies, rather than RCTs, may contribute to earlier patient access. Specifically, as illustrated schematically in Figure 1 and described in section 1.5.2, approval will happen earlier when pivotal evidence comes from uncontrolled phase II trials measuring surrogate endpoints, i.e. single-arm, interventional non-randomised studies. Regulators are also considering an expanded role for non-interventional non-randomised (i.e. observational) studies in providing evidence for regulatory approval (see sections 2.4.3 and 2.4.4). Approval based on non-randomised studies is more commonly associated with expedited approval pathways, but, as demonstrated in the literature (sections 2.4.1 and 2.4.2) and in two empirical studies presented in this thesis (chapters 4 and 5), also occurs under regular approval provisions.

While the use of non-randomised studies may accelerate the time until new medicines are available on the market, the benefits and risks of using less methodologically robust data need to be considered. As discussed in the following sub-sections, the advantages and drawbacks of approval based on more uncertain evidence may be different for medicines developers, patients, and other stakeholders.

6.2.1. Reduced clinical development times and costs

Expedited approval pathways and approval based on non-randomised studies represent efforts to speed up market access for new medicines. Indeed, expedited approval pathways have been shown to significantly shorten the time until marketing authorisation. In the US, approval via expedited pathways was associated with a 1.2-year shorter clinical development time compared to regular approval.[300] This was particularly pronounced for the accelerated approval pathway, which allows approval using early evidence, such as single-arm phase II trials. For cancer medicines, considerably shorter clinical development times were found for expedited pathways in both the US and Europe, and these were shortest for conditional approval pathways (i.e. accelerated approval in the US and conditional marketing authorisation in the EU).[301]

Because they result in earlier access to new medicines, shorter development times may be beneficial for patients if these medicines prove to be effective. Whether this is indeed the case has been questioned by empirical evidence about the proven patient benefit of medicines approved through expedited pathways and / or based on non-randomised studies, including data presented in this thesis, (see section 6.2.2 below). For medicines developers, on the other hand, the benefits of shorter development times may exist even without proven clinical value. To consider the benefits of approval based on early evidence, it is instructive to revisit Figure 1 in the introduction section and consider the bottom panel of that figure, which illustrates the monopoly protection period for new medicines. Bringing market approval forward (e.g. through authorisation based on non-randomised phase II data) also results in an earlier start of the effective monopoly period during which companies are protected from downward pressure on prices through generic or biosimilar versions of their product.

While the mechanism for generating revenue through early approvals is clear, the empirical evidence suggests a more complex relationship between time of market

access, evidence robustness, and revenue generation. Early approval also means less comprehensive data on clinical benefit. A study of cancer medicines approved by the EMA from 1995 to 2020 compared their added benefit ratings as assessed by HTA bodies and professional societies with the revenues generated by these products.[302] Revenues tended to be higher for products with better added benefit ratings, although that relationship was not statistically significant. The study also found that medicines with regular approval generated more revenue compared to those with conditional marketing authorisation. This may reflect the more comprehensive evidence package available at the time of approval. However, lower total revenues for conditionally approved medicines may also be explained by the fact that these products are more likely to be for orphan indications, where market size is smaller.

Should we therefore expect medicines developers to invest in more robust evidence to generate higher revenues? This is a largely hypothetical question about a counterfactual that is unlikely to materialise. For most medicines approved based on limited data, robust confirmatory evidence on patient-relevant outcomes that may result in a higher added benefit rating never emerges. This pattern was shown in chapters 4 and 5 of this thesis and in several previous studies on cancer medicines approvals from the US (see section 2.5.2).[61, 142, 155] In practice, the ability to generate any revenue depends on receiving approval first, and for products approved based on limited evidence, more robust evidence may never emerge.

Beyond the earlier start of revenue generation, pharmaceutical companies benefit more directly from approval based on less robust trials due to their lower cost. In a review of clinical trials used to support FDA approval of 59 new medicines from 2015 to 2016, the average cost of single-arm trials was less than half that of trials with comparators.[303] Trial costs were also lower for trials measuring surrogate endpoints, rather than clinical outcomes.

In summary, medicines developers are overall expected to benefit from approval based on less complete data, which allows early revenue generation at lower cost compared to approval based on comprehensive evidence packages which take longer to generate. The benefits to patients, on the other hand, are not as clear, as discussed in section 6.2.2 and 6.2.3 below.

6.2.2. Clinical benefits to patients

A strand of literature has assessed the clinical benefit of newly authorised medicines to patients, including medicines approved through expedited pathways and on the basis of more uncertain evidence. While regulatory approval determines whether the benefits of the product are likely to outweigh its risks, HTA bodies, health care payers, and physicians, including their professional societies, are interested in the added benefit compared to existing alternatives. Evidence on added benefit is needed for these actors to inform decisions on what treatments to prescribe and reimburse.[270] Studies focusing on clinical benefit of medicines approved through expedited pathways or on the basis of more uncertain evidence have overall found a minority of these medicines to provide added therapeutic value.

Clinical benefit of products with limited evidence

Between 38% (for EMA approvals) and 39% (for FDA approvals) of products with conditional approval were shown to have high added therapeutic value.[190] This was considerably lower for cancer medicines compared to other indications (31% and 36% for cancer products in the EU and US, respectively, vs. approximately 53% for non-cancer products in both markets). For Europe, Brinkhuis et al. also showed that the majority (57%) of cancer medicines with conditional marketing authorisation from 1995 to 2020 had a negative or non-quantifiable added benefit rating, and this proportion was significantly higher than for standard approvals.[302] Beyond conditional approval pathways, which explicitly allow more limited evidence for initial approval, the evidence on the relationship between expedited approval pathways and added benefit is mixed. Only about half of medicines submitted through the priority review (FDA) and accelerated assessment (EMA) pathways between 2006 and 2015 were of high clinical added value (42% for FDA approved drugs, 54% for EMA approved drugs).[232] Another study found that some of the FDA's expedited pathways (i.e. priority review, fast track designation, and breakthrough therapy designation) – but not accelerated approval – were associated with higher added benefit ratings compared to approvals without any expedited programme.[189] However, the majority of these medicines were still rated as providing low added benefit. While these pathways are not directly linked to evidence standards for approval, the relatively small proportion of products with

high added benefit rating is remarkable since these expedited pathways are intended for products promising substantial therapeutic innovation.

Similarly to studies on expedited approvals, the less extensive literature on products authorised based on non-randomised studies shows that high added benefit ratings for these medicines are rare. For cancer medicines approved by the FDA from 2006 to 2016, Tibau et al. found that 7% of indications with single-arm trial data had a substantial clinical benefit as indicated by the ESMO-MCBS scale compared to 41% of indications with RCT evidence.[304] For Europe, Mulder et al. reported that only three of 21 single-arm trials used as pivotal evidence for conditional EMA approval of cancer medicines from 2012 to 2021 were assigned an ESMO-MCBS score indicating substantial added benefit.[29]

In addition to added benefit assessments, demonstrated clinical benefit has also been the subject of a body of literature focusing on confirmatory post-marketing studies. This strand of literature was described in the literature review (section 2.5.2). Summarised briefly, empirical studies have shown that confirmation of clinical benefit on patient-relevant outcomes based on robustly designed studies only emerges for a minority of medicines first approved based on limited evidence. For cancer medicines with accelerated approval in the US, RCTs showing overall survival benefit have only emerged for approximately one fifth of products.[61, 142, 155] Adding to this body of literature, novel findings presented in this thesis show similar figures for confirmation of overall survival benefit after approval based on single-arm trial data (chapter 5).

Methodological issues and their impact on added benefit ratings

The low number of high added benefit ratings for products approved through expedited pathways may partly be explained by the type of evidence available at the time of approval. For conditional approvals in particular, this is expected to be less complete and methodologically robust than for traditional approval. The methodological features of early approval, such as non-randomised studies and use of surrogate endpoints instead of overall survival and quality of life, may directly impact on added benefit ratings, as HTA bodies and professional societies take the quality of evidence into consideration when rating the therapeutic value of medicines.[305]

A publication by the German HTA body, IQWiG, highlights how methodological weaknesses in clinical evidence impact on added benefit assessment.[276] In an analysis

of all 216 medicines entering the German market from 2011 to 2017, only 25% were assigned a major benefit rating, which requires demonstration of a significant effect compared to the standard of care on a patient-relevant outcome in a RCT or a very large effect in a non-randomised study. For 7% of products, the added benefit was non-quantifiable (e.g. due to lack of a relevant comparator), and for the majority (58%), no added benefit was demonstrated – the main reason being lack of a study comparing the new medicine with the standard of care. A lack of added benefit rating may therefore be explained by the absence of relevant evidence. This includes cases where robust evidence from a RCT is available, but not for the comparator of interest, but may also include cases where only evidence from non-randomised studies with methodological issues is available. In any case, the substantial proportion of lack of added benefit ratings due to limitations in the evidence base reveals a key issue of regulatory approval based on limited data, such as non-randomised studies: in the absence of methodologically robust evidence, added benefit may either not be quantifiable or judged to be inexistent. The uncertainty about clinical benefit for medicines approved based on non-randomised studies or other early evidence can therefore result in negative reimbursement decisions and lack of patient access.

6.2.3. Implications of uncertain evidence on downstream decision making

Whether and to what extent patients benefit from medicines authorised based on methodologically less robust evidence is not only determined by a positive regulatory outcome, but also by availability and take-up of these products, i.e. even when a product is deemed safe and effective at regulatory approval, it can only have a meaningful impact on patients' lives if it used. As briefly introduced in section 1.5, in many health systems, particularly in Europe, access to new medicines is contingent upon a positive reimbursement recommendation or decision by HTA bodies and / or payers. These actors consider the added value of a new medicine in relation to its cost and existing alternatives (in principle, individual patients could also pay for medicines privately, thus bypassing the reimbursement decision of payers, but this is considered less relevant for newly authorised medicines which enter the market at increasingly high prices). Finally, physicians must be convinced about the benefit of the product, and patients must be willing to take them. These actors, collectively referred to as

downstream decision makers, require information about the added benefit of newly authorised medicines.[270]

As indicated in section 6.2.2 above, the methodological weaknesses of non-randomised studies and early evidence submitted under expedited pathways pose challenges for determining the added benefit. The empirical literature shows that these challenges may preclude medicines authorised based on less complete evidence than usual from being available to patients earlier, and in some cases from being available at all. Evidence presented in this thesis has underlined the importance of the uncertainty in clinical evidence at the time of approval, as non-randomised studies may generate different effect estimates compared to RCTs (chapter 3), and uncertainty due to weak study designs may persist even years after initial approval (chapters 4 and 5). Evidence at the time of initial approval therefore often represents the most robust evidence available, and downstream decision makers have to rely on that same evidence.

Time to patient access

Contrary to its intentions, early regulatory approval may not lead to faster patient access because downstream actors face challenges in making decisions based on uncertain evidence. For products approved based on early evidence, such as single-arm trial data, the lack of comparative data limits the assessment of a product's added value. While there are methods for comparing data from single-arm trials to external controls (see also section 2.4.2), these are fraught with uncertainties.

In Europe, where national or regional processes for determining pricing and coverage of medicines through national health systems or social health insurance exist, expedited approval pathways are not associated with shorter time until patient access. A Belgian study found that expedited approval pathways at the EMA were overall not associated with reduced time until the products became available in Belgium; while pathways targeting the review time (i.e. accelerated assessment) and interactions between regulators and medicines developers (i.e. PRIME) were associated with shorter regulatory review times, this did not result in reduced completion times for the national pricing and reimbursement process.[306] Pathways allowing approval based on less complete evidence than normally required (i.e. conditional marketing authorisation and authorisation under exceptional circumstances) were not associated with reduced regulatory review times, and also did not result in shorter time until completion of the

national process compared to standard approvals. In some jurisdictions, conditional approval may even lead to longer time until patient access. A study of HTA outcomes across four countries (Canada, England, France, and Scotland) found that conditional approval was associated with a delay of approximately six months until favourable reimbursement recommendation compared to standard approval.[307]

Coverage decisions based on uncertain evidence

A possible explanation for the lack of early availability of medicines approved through expedited pathways is that the actors involved in pricing and reimbursement decisions require time to prepare and review evidence packages based on less methodologically robust data. This includes HTA bodies and/or payers conducting assessments and making decisions, as well as pharmaceutical companies submitting evidence, including data to resolve uncertainties raised by authorities. Such remaining uncertainties have been documented to be more prevalent for medicines authorised through a conditional approval pathway compared to standard approval.[307] Indeed, addressing open questions during the pricing and reimbursement process takes time. Data from Italy show that the net time for HTA and pricing negotiations for on-patent products (excluding time for administrative checks, final decision and publication) was approximately nine months.[308]

Evidence from non-randomised studies presents a particular challenge for determining added benefit. While regulatory agencies may deem evidence from single-arm trials sufficient to establish that a causal treatment effect of the new medicine exists (e.g. for cancer medicines, documented tumour response for a condition with well-established natural course of disease), HTA bodies and payers prefer evidence on patient-relevant outcomes such as survival or quality of life.[305] However, such data requires contextualisation to determine whether the outcome observed for the novel medicine represents an added benefit compared to existing treatments. Accordingly, HTA bodies and payers have traditionally valued RCT evidence even higher than medicines regulatory bodies, and lack of methodologically robust evidence at time of approval has been shown to be associated with lower ratings of added therapeutic value by these bodies (see section 6.2.2 above). This is consistent with a more conservative approach to accepting uncertainty: as underscored by the empirical evidence presented in this thesis (chapter 3), treatment effects obtained from non-randomised studies carry

uncertainty and may lead to different conclusions about whether a treatment works for a substantial proportion of clinical questions.

Uncertainty in the clinical evidence therefore presents a barrier to reimbursement of new medicines and may ultimately limit their availability to patients. When faced with limited evidence on a medicine's effectiveness, HTA bodies and payers may decide against a recommendation to reimburse the product (i.e. reject reimbursement) or recommend reimbursement with restrictions (also referred to as listing with restrictions in the literature). The literature shows that limited clinical evidence is indeed associated with either negative or restricted reimbursement decisions. In a study covering 40 countries, less than half (48%) of HTA submissions based on single-arm trial data obtained a positive or restricted positive recommendation, compared to approximately two thirds of all HTA submissions.[120] Other studies found that HTA bodies were more likely to issue a negative recommendation for medicine indications with conditional approval compared to standard approval and that unrestricted positive recommendations were overall rare for these products.[307, 309] Interestingly, one of these studies found that, within the group of products with limited evidence as indicated by conditional approval, the study design for generating clinical evidence was not independently associated with more positive or negative HTA recommendations, as there was no statistically significant difference in HTA outcomes for conditionally approved products with controlled vs. uncontrolled clinical trial data.[309] However, medicines with important methodological weaknesses may not even be submitted for consideration for reimbursement in some health systems, as was shown for medicines with EMA approval based on non-randomised studies.[167] These medicines may therefore not become available as standard treatment option in all settings.

To summarise, it has been empirically shown that a minority of medicines approved through expedited pathways or based on non-randomised studies provide a meaningful therapeutic benefit over existing treatments. Moreover, approval with methodologically limited evidence presents challenges for making reimbursement decisions, resulting in overall more limited – rather than expedited – availability of medicines for patients. The uncertainty about the clinical benefit of these medicines may mean that no relevant added benefit over existing alternatives exist, in which case no major benefit to patients would be expected, or it means that the added benefit

cannot be determined to make the treatment available to all patients. Either way, the uncertainty introduced by early regulatory approval based on non-randomised studies or through expedited pathways raises questions about their contributions to population health.

6.3. Regulatory approval as signal for evidence standards

The evidence standards set by medicines regulatory bodies are instrumental in shaping the body of evidence available for new medicines. Research presented in this thesis, as well as other studies, demonstrate that the evidence generated to support initial marketing authorisation is often the only (robust) evidence on the clinical benefit of new medicines in the authorised indication. The standards applied at the point of approval are therefore highly consequential: once a medicine is authorised, further robust evidence may not become available for that same indication, and uncertainty about the clinical benefit in the patient group for which approval was first granted may never be fully addressed. Even when RCT evidence for the same indication becomes available (as was shown to be the case for approximately half of cancer medicine indications in chapter 5), it takes time to emerge and may still not address open questions about patient-relevant outcomes, and initial treatment and funding decision are therefore subject to the uncertainty of methodologically limited evidence.

Common exceptions

Expedited approval pathways were introduced to manage the trade-offs between uncertainty and access to new medicines in areas of unmet need: in exceptional cases (when the disease is severe and no other treatment exists), early evidence can be acceptable to authorise market entry of a new product. While expedited approval pathways do not change the fact that new products still need to be deemed to provide more benefit than harm, the evidence to support such claim is accepted to be less robust. However, as shown in this thesis and other studies, approval based on early evidence (often from single-arm trials) is not necessarily exceptional. For cancer, a substantial proportion of new medicines (up to half of all new cancer medicines approved in the EU, see chapter 5) are now approved based on data from single-arm trials, without results from RCTs available. Moreover, these approvals do not always

occur through expedited pathways but include regular approvals. These findings indicate a shift towards increasing regulatory acceptance of uncertainty.

It is important to consider the consequences of this trajectory: as regulatory acceptance of less methodologically robust evidence increases, the evidence base for informing treatment decisions (for patients and physicians) and for funding access to new medicines (for HTA bodies and payers) erodes. Comparative benefit assessment to decide which of two (or more) treatments to pursue for individual patients, and which of these to cover in a health system, relies on methodologically robust evidence (see section 6.2.2). Ideally, these would be head-to-head RCTs comparing the new treatment to the standard of care.[270] Even when no direct evidence comparing one medicine against another is available, methodologically robust evidence on that medicine against another can be used to infer comparative effectiveness through network meta-analysis.[310] However, this relies on the availability of well-conducted RCTs for the indication of interest. As the exception of regulatory approval based on methodologically limited evidence, including single-arm trials, becomes the norm in a therapeutic area, comparative benefit assessments are increasingly jeopardised, leaving downstream decision makers to make treatment and funding decisions under considerable uncertainty.

Impact of regulatory standards on evidence generation

The threshold for regulatory acceptance of uncertainty sends an important signal to medicine developers. As discussed in section 6.2.1, generating more robust evidence costs time and money. If regulators are willing to accept comparatively weak evidence for initial approval, there are no incentives for pharmaceutical companies to invest in more robust data. This matters not only for the evidence available for regulatory decisions but also creates challenges for determining the added benefit of new medicines, as discussed in section 6.2.2. While regulatory approval and HTA differ in scope, HTA bodies often have to inform pricing and reimbursement decisions with the same evidence submitted for regulatory approval. Evidence standards that impede comparative benefit assessments, such as acceptance of single-arm trials or comparisons against inferior treatment alternatives, therefore contribute to uncertainty about a new medicine's place in therapy.

While regulators may emphasise the scope of their assessment (i.e. focusing on the benefit-risk ratio, but not the comparative added therapeutic value), the standards they adopt have consequences beyond the regulatory domain. As signals to medicine developers, regulatory evidence standards shape how development programmes are designed and directly influence whether industry prioritises the generation of robust comparative evidence. Robust evidence requirements for regulatory approval therefore serve a dual purpose: they protect public health in the immediate term but also create incentives for industry to develop better medicines.

Statutory vs. revealed preferences

An important consideration, then, is how regulators convey their standards for approval. In their guidance documents, regulatory bodies maintain traditional evidence standards. For example, ICH guidance on the choice of control groups adopted by both EMA and FDA clearly highlights the methodological advantages of RCTs, showing a preference for this study design whenever possible.[16, 162] The EMA also indicates a preference for RCTs in other guidance documents, including in a reflection paper on the use of single-arm trials.[17, 23] Yet, empirical analyses presented in this thesis, as well as research conducted by others, shows that regulatory bodies employ considerable flexibility when interpreting these standards. In oncology, a large proportion of new medicines approvals are now based on single-arm trials, rather than RCTs. This flexibility is not out of line with guidance documents. Both EMA and FDA describe situations where RCTs may not be feasible or necessary, in particular in relation to situations of unmet need and when the natural history of disease is sufficiently well-known, so that a biological response can be ascribed to a treatment even without concurrent control.[18, 23] Nevertheless, the increasing acceptance of non-randomised evidence for regulatory approval in some therapeutic areas may indicate a change in effective evidence standards.

This potential shift in effective evidence standards is further reinforced by regulatory initiatives to expand the use of non-randomised evidence. While guidance documents continue to demonstrate a preference for RCTs, regulatory bodies have in parallel set actions to promote the use of real-world evidence.[35, 39] These include a potential role for non-randomised studies not only to address uncertainties in the post-marketing setting but to also support initial marketing authorisations. These initiatives

send signals to the pharmaceutical industry that, despite its methodological limitations, non-randomised evidence is expected to play an increasingly important role for regulatory approval.

Taken together, the revealed preferences of regulatory bodies as demonstrated through approval decisions and frameworks may contribute to the promotion of non-randomised evidence. Rather than promoting the conduct of RCTs by improving their feasibility even in challenging circumstances, the increasing reliance on non-randomised evidence may in itself lead to further promotion of non-randomised studies.[227] This trend increasingly challenges comparative benefit assessments and results in uncertainty when making treatment and funding decisions.

6.4. Policy implications

The findings presented in this thesis have important implications for regulatory policy, evidence generation, and ultimately patient access to new medicines. While regulatory approval is often framed as balancing timely access with robust evidence standards, the empirical studies in this thesis highlight the important role of regulatory approval in shaping evidence generation, impacting on the entire pharmaceutical value chain. The empirical studies of regulatory decisions presented in chapters 4 and 5 focused on cancer as an illustrative example of evidence standards and the management of uncertainty, but given the common use of expedited approval pathways for non-cancer medicines,[300] including some of the same methodological limitations in the clinical evidence in relation to the use of non-randomised studies and non-validated surrogate endpoints,[144, 311, 312] the policy implications derived from the thesis are relevant across therapeutic areas. Key policy implications about evidence generation from a pharmaceutical value chain perspective, selective use of expedited approval pathways, and transparency about regulatory acceptance of uncertainty are outlined in the subsections below.

6.4.1. Pharmaceutical value chain perspective on uncertainty about clinical benefits

As discussed in section 6.2.3, the evidence accepted for regulatory approval has implications for downstream decision making. Uncertainty about the clinical benefit

due to non-randomised pivotal trials impacts on assessments for determining the new medicine's place in therapy. From a regulatory perspective, the decision to grant approval may be accompanied by regulatory measures described in chapter 4, including conditional approval, post-marketing obligations, and limiting the authorised indication. For regulators, these measures may sufficiently address remaining uncertainties to allow the product to enter the market. For other actors, the uncertainties in the evidence base at the time of approval are perceived as more substantial. For HTA bodies and payers, the evidence may be insufficient to inform comparative effectiveness assessments, creating difficulties for reimbursement decisions. For physicians, uncertainties about the clinical value of newly authorised medicines pose a challenge when integrating them into treatment pathways. For patients, uncertainty about the therapeutic value means they may be exposed to treatments without proven therapeutic benefit.

Regulatory approval as key to availability of robust evidence

Addressing uncertainty stemming from methodologically weak clinical evidence should therefore be considered as key challenge across the pharmaceutical value chain. A holistic approach for addressing the evidence needs of all decision makers is needed. As discussed in 6.2.3, this includes robust evidence on the comparative effectiveness of the new medicine relative to the standard of care. The focus of evidence generation should therefore not only rest on more methodologically robust evidence, i.e. well-designed and conducted RCTs that are transparently reported. The evidence should allow a meaningful assessment of the added therapeutic value compared to alternative treatments, i.e. there is a requirement for RCTs with active comparators to provide direct evidence on comparative effectiveness.[270]

Within a pharmaceutical value chain approach to evidence generation, the role of regulatory evidence standards remains key: since regulatory approval represents the first gateway on the path to patient access, the standards applied at this stage shape the evidence generation programmes for new medicines. Regulatory acceptance of methodologically weaker evidence signals to medicines developers that similar evidence will be sufficient in the future (see also the discussion on regulatory approval as signal for evidence standards in section 6.3). It is therefore important for regulators to occupy this role pro-actively and demand robust evidence, rather than seek solutions

to accommodate weaker evidence. The current trend towards increasing regulatory approval on the basis of non-randomised evidence (at least in some therapeutic areas) may create a cycle of uncertainty in which comparative effectiveness cannot be reliably established because only evidence from uncontrolled studies is available.

The contrast between the value chain perspective on evidence generation, with a focus on active-controlled RCTs, and current regulatory practices, with increasing acceptance of uncertainty at the time of initial approval, may be construed as a dichotomous choice in how regulators position themselves: regulators could be perceived acting primarily as guardians of public health, with a cautious approach to evidence generation that prioritises certainty, or as promoters of pharmaceutical innovation, embracing novel forms of evidence and a life cycle approach to evidence generation. However, such dichotomy poses a false trade-off between evidence standards and innovation. Regulatory acceptance of uncertainty may reduce incentives for pharmaceutical companies to invest in innovative products providing substantial added therapeutic value. Without the need to generate robust evidence on added benefit, medicines developers can achieve the same regulatory outcome with more limited evidence. Rather than focusing on R&D programmes for breakthrough therapies that seek to establish superior therapeutic benefit – but may fail to demonstrate such added value – companies may rationally prioritise methodologically weaker evidence which makes it more difficult to establish the added benefit. Robust evidence requirements, on the other hand, may incentivise a focus on medicines with substantial therapeutic value. By maintaining high evidence standards, regulators can therefore fulfil both roles simultaneously: ensuring that only products which demonstrably improve patient outcomes enter the market creates incentives for developing innovative medicines that provide real therapeutic value.

Robust evidence standards therefore protect public health while fostering an ecosystem where new medicines provide an added therapeutic benefit which is recognised (through substantial evidence) by HTA bodies, payers, and clinicians, and leads to faster uptake of these products in clinical practice, ultimately benefiting patients.

Aligning regulatory and HTA perspectives

What would a pharmaceutical value chain perspective on evidence generation look like in practice? As described above, the evidence requirements for reimbursement and pricing decisions, as well as for informing treatment decisions in clinical practice, are different than for regulatory approval but due to its position as first gatekeeper for new medicines entering the market, the evidence standards for the latter are shaping R&D programmes. An alignment is therefore needed between regulatory standards and the requirements of downstream decision makers.

Specifically, there is scope for aligning the evidence requirements of regulators with HTA bodies and payers. While the idea of breaking down silos in the pharmaceutical value chain and bringing together regulators with HTA bodies and payers is not new,[313] recent policy developments in Europe have increased the chances of turning this into reality. As part of wide-ranging reforms of the EU's pharmaceutical legislation affecting both regulatory approval (still under negotiation at the time of writing of this thesis) and HTA (implementing EU-wide joint clinical assessments / JCAs which need to be taken into consideration for national-level HTA, starting from 2025), joint scientific consultations (JSCs) are foreseen to guide the evidence generation programmes of medicines developers. In addition to the separate processes of scientific advice by EMA and JSC by the coordination group of HTA bodies, developers may also request a parallel JSC with both regulators and HTA bodies.[314] For pharmaceutical companies, this instrument may prove helpful in clarifying evidence requirements for different purposes early on and shaping evidence generation plans accordingly. It is unclear whether companies would adhere to the advice given – prior experience suggests 75% compliance with scientific advice on evidence generation by EMA.[315] However, there is reason to believe that parallel JSCs could have an impact on companies' development programmes, since the implementation of EU-wide joint clinical assessments means that advice on HTA evidence requirements now account for the entire European market, a very different situation from previous joint consultations in fragmented health systems in the EU.

Parallel JSCs may therefore reduce divergence in evidentiary standards and ensure that data generated for regulatory approval is also fit for reimbursement decisions. Such alignment could therefore both strengthen evidence standards and improve patient

access to effective medicines. Another route for aligning evidence standards is to leverage the power of payers and turn their evidence requirements into the most important gateway for market access.[270] If payers do not accept methodologically weak evidence, or if they tie prices to the availability of robust evidence demonstrating added value, pharmaceutical companies may react by shaping their evidence generation plans according to the needs of payers. However, given the current evidence landscape for newly authorised medicines – in particular in cancer – this would effectively result in a lack of patient access to these new medicines, either because of negative reimbursement decisions due to lack of proven added benefit, or because companies would not market their products at low prices reflecting the lack of robust evidence.

Strengthening evidence requirements at the regulatory level by aligning them with the needs of HTA bodies and payers is expected to result in improved patient access to new medicines, contingent on their proven added benefit. Such aligned evidence standards would directly address the open questions regarding the patient benefit of regulatory approval based on limited evidence described in section 6.2.2 since newly authorised products would have the required evidence to inform comparative effectiveness analysis and to support claims for reimbursement and their place in therapy. The impacts of a potentially more holistic perspective on evidence generation remain uncertain for now. However, JSCs and the introduction of joint clinical assessments across the EU will provide an opportunity to study whether there is appetite for such alignment from regulators, HTA bodies, and – importantly – the pharmaceutical industry, and, if so, whether early engagement between medicines developers and different public authorities result in more robust evidence generation programmes, and ultimately in patient access to new medicines with proven therapeutic benefit.

6.4.2. Selective use of expedited approval pathways

Expedited approval pathways have become an increasingly prominent feature of regulatory policy in oncology and other therapeutic areas.[300] These pathways were intended to accelerate access to innovative medicines addressing unmet medical needs in life-threatening or seriously debilitating conditions. However, the evidence presented in this thesis (focusing on cancer) and by previous research raises concerns

regarding the robustness of the cumulative evidence across pre- and post-marketing phases for products approved under expedited pathways and how targeted these pathways are currently used. A more selective approach to granting medicines the benefits of expedited approval may be warranted.

Ensuring robust confirmatory evidence

From a methodological point of view, non-randomised studies play an important role in expedited pathways (see chapter 4, and also section 2.3 for a literature review on the association of expedited pathways with less robust evidence, including different types of non-randomised studies). For conditional approval, this potentially includes initial approval based on a single-arm trial followed by a confirmatory RCT, but – as documented for EMA approvals in cancer in particular (chapter 4) may also include confirmatory evidence from a single-arm trial and other post-marketing evidence generated through non-randomised studies, such as observational studies. Even for approvals based on RCT evidence, confirmatory evidence through non-randomised studies is envisioned as the importance of real-world evidence for regulatory decisions increases (see section 2.4.4). Historically, conditional approval typically included a confirmatory RCT, but the empirical evidence suggests that the cumulative evidence packages available for medicines approved through expedited pathways do not address all uncertainties about their clinical benefit. This is partly because the methodological limitations of pivotal trials, including non-randomised designs and surrogate endpoints, are mirrored in confirmatory post-marketing studies.

It is therefore important to design confirmatory studies as robustly as possible to ensure that remaining uncertainties are fully addressed. In practice, this would mean RCTs with active controls measuring patient-relevant outcomes whenever these are feasible. Such RCTs provide the evidence required to demonstrate the added therapeutic benefit compared to the standard of care. As discussed in chapter 5, there are valid concerns regarding the feasibility and ethics of running RCTs in situations where existing treatment options are limited and patient numbers are low, resulting in potentially overly long clinical trial duration and delays in access to the new medicine. While RCTs may not be possible to conduct for all indications, their benefits in terms of reducing uncertainty about the therapeutic value of new medicines justify increased

efforts to improve their feasibility, both to provide pivotal evidence but also as confirmatory trials when early evidence justifies expedited approval.

Improving the feasibility of RCTs

The lack of feasibility and the ethical concerns about running RCTs present real challenges for regulatory approval of new medicines. However, the situations where these concerns represent true barriers to generating robust evidence through RCTs may be more limited than previously thought. The empirical evidence presented in chapter 5 of this thesis suggests that, for many cancer medicine indications, RCTs may not require larger sample sizes and do not take longer to complete than pivotal single-arm trials. While RCTs are still unlikely to be feasible for all indications, regulators should work with medicines developers and other stakeholders in the health system to strive for the most robust evidence possible. Just as substantial resources have been committed by public and private actors to develop methods and standards for using observational (real-world) data for regulatory decisions, more efforts may be required to render RCT easier and cheaper to conduct, so that they can provide relevant evidence in a more timely fashion.[227]

There are measures that can be taken to improve the feasibility of RCTs and to ensure that they provide the evidence required, including when intended as confirmatory study after early approval based on more limited evidence. Importantly, post-marketing evidence generation needs to be set up well before initial approval. Chapter 4 of this thesis (focusing on cancer medicines) and a range of other studies [61, 62, 74, 185] have documented delays in post-marketing studies, including studies intended to confirm clinical benefit of products with conditional approval. Timely confirmation of clinical benefit is key to the concept of conditional approval: without the promise of forthcoming confirmatory evidence, conditional approval would represent an effective lowering of the evidence standards for regulatory approval. A study from the US sheds some light on the importance of timely initiation of post-marketing studies: among all medicines with accelerated approval from 2009 to 2019, confirmatory studies were underway at time of approval for 72% and were not underway for 20% (for the remainder, no information was available).[316] Studies that were underway at time of approval were less likely to be delayed compared to studies that were not underway, suggesting that early planning and initiation of studies

contributes to whether or not confirmatory evidence becomes available in a timely manner. Such early planning of methodologically robust confirmatory studies should feature in the parallel scientific advice provided by regulators and downstream decision makers (see section 6.4.1 above).

Another measure to improve the feasibility of RCTs is to integrate them more effectively into routine care, and to leverage international collaboration. Pragmatic RCTs which reduce the administrative burden have been conducted in different therapeutic areas, including for COVID-19 and myocardial infarction.[227] For rare diseases, RCTs may be easier to conduct within disease registries when joining forces internationally, increasing patient numbers to recruit the required sample size. Regulatory guidance on conducting RCTs in routine care may further increase interest in optimising trial designs to improve feasibility. Aligning such guidance between regulators and downstream decision makers would likely magnify its impact.

Expedited pathways may include early approval based on limited evidence. In these cases, robust confirmatory evidence should be required. While concerns exist regarding the feasibility of RCTs, expedited pathways may provide an opportunity for regulators and HTA bodies to engage with medicines developers and convey evidence requirements early. Rather than finding ways to accommodate weaker evidence, these pathways should be used to incentivise methodologically robust confirmatory evidence.

Innovativeness of products approved through expedited pathways

There are also questions about how targeted expedited approval pathways are towards innovative products. In the US, the use of expedited approval pathways has increased significantly over time, leading to a situation where the exception (i.e. approval through one of the FDA's expedited pathways) has become the norm: about two thirds of newly approved medicines in 2013 were reviewed under accelerated approval, priority review, fast track or orphan disease designation.[55] This development was driven by non-first in class drugs (i.e. by medicines that are based on existing chemical compounds), calling into question the innovativeness of products with expedited market access. By 2018, the proportion of FDA approvals through expedited pathways had increased to three quarters.[270] At the same time, the evidence suggests that most of these products do not represent a true therapeutic advantage.

Overall, a more selective approach to expedited approvals is needed to align exceptional regulatory acceptance of uncertainty with exceptional patient benefits. Identifying products with such promise is clearly challenging, but the current approach suggests a lack of discrimination between true breakthrough products and other medicines which do not offer major added benefit. When considering candidates for expedited approval, the evidence generation programme should play a role to ensure that robust evidence demonstrating the promised substantial added benefit is forthcoming.

6.4.3. Transparency about regulatory acceptance of uncertainty

Finally, greater transparency is needed about regulatory acceptance of uncertainty, and in particular acceptance of non-randomised studies for initial approval. As shown in chapter 5 of this thesis for cancer medicines, justification for acceptance of single-arm trials for regulatory approval is often lacking in publicly available documents. This appears to be in stark contrast to regulatory guidance which emphasises the importance of methodologically rigorous studies for approval and for deviations from RCTs to be justified. While regulators commonly discuss limitations of submitted evidence and explicitly list uncertainties in relation to the benefit-risk profile, some of which may be due to the design of pivotal trials, the reasoning behind acceptance of such limited evidence is often not communicated clearly. Addressing the obvious deviation from the stated gold standard for regulatory approval directly is necessary to better communicate the trade-offs made by regulators. Such open discussion of justifications for regulatory acceptance of uncertainty would also allow a broader dialogue about evidence standards and what they mean for access to effective medicines.

Against the novel evidence on feasibility of RCTs for cancer medicines approved based on single-arm trials presented in this thesis, transparency about regulatory acceptance of non-randomised evidence is particularly important. It is important to acknowledge that RCTs may not always be feasible or ethical to conduct and that there is scope for generating pivotal evidence through single-arm trials. However, for a substantial proportion of cancer medicines approved by the EMA from 2014 to 2023 based on single-arm trial data, RCTs in the same patient population as the initially approved indication were identified (chapter 5); furthermore, the study indicated that

commonly stated reasons for the lack of feasibility of RCTs – namely, a small patient population and the longer time required to run a RCT – may not apply to approximately half of these indications. Given the limitations for obtaining causal effect estimates from these studies, scholars have recommend establishing a framework for when evidence from single-arm trials is acceptable for regulatory approval, emphasising the importance of clearly communicating the methodological shortcomings of approvals based on uncontrolled studies.[76] From a study design perspective, the recommendations focus on appropriate historical controls which could come from multiple sources to ensure close matches with the participants of the pivotal single-arm trial, or alternatively, the pivotal trial should employ inclusion and exclusion criteria to closely match existing historical cohorts. The uncertainties due to the uncontrolled nature of the pivotal trial should, at the latest, be addressed through confirmatory RCTs.[76] The methodological recommendations are also reflected in general design principles for pivotal trials in regulatory guidance.[16, 17, 162] However, these principles do not appear to be consistently adhere, as shown by previous studies and the research presented in this thesis: firstly, external controls are often only included as implicit benchmark, and when used as direct comparator, they are often not adjusted, and secondly, confirmatory RCTs may not emerge at all, and when they do, may not confirm added benefit.

Against this background, more transparency about the uncertainty deemed acceptable for regulators is required. For approvals based on single-arm trials, medicines developers should clearly explain why they were not able to conduct a RCT, and regulators should discuss the validity of these explanations in publicly available documents. Regulators should further explicitly state the expected benefits and risks of foregoing RCT evidence for initial approval. As discussed in sections 6.3 and 6.4.1 above, applying robust evidence standards may address the dual objectives of medicines regulators, i.e. to ensure only safe and effective medicines enter the market and to promote innovation by focusing on treatments that provide significant added benefit. An informed and evidence-based dialogue with patients, the public, and academics about the perceived trade-offs between timely access to new medicines and robust evidence standards could help clarify the role of regulators in the 21st century.

6.5. Limitations of this thesis

Limitations specific to the three studies in this thesis were described in the respective chapters (sections 3.4.5, 4.4.8, and 5.4.5, respectively). Beyond these, there are some limitations that are relevant across the thesis, and these are discussed below.

Firstly, there are limits to the generalisability of the findings of empirical work presented in this thesis. This thesis analyses regulatory approvals in Europe and the US, with a particular focus on cancer medicines approved either through specific regulatory pathways or based on non-randomised evidence. The regulatory bodies in these two settings, EMA and FDA, respectively, are considered globally leading in regulatory science. Focusing on their evidence standards and how they manage uncertainty associated with non-randomised studies may therefore be instructive, as their frameworks and approaches may be followed by other regulators. However, other agencies may also develop their own frameworks for non-randomised evidence. Importantly, China is rapidly emerging as a key market for new medicines, with an increasing number of products first launched in China.[317]

The focus on cancer medicines is justified by the frequency of expedited approvals and approvals based on non-randomised studies in this therapeutic area, but findings regarding appropriate study designs (in particular in relation to response rate as endpoint in single-arm trials) and the relevance of non-randomised evidence for expedited and regular approvals may not be transferable to other disease areas. There are therefore limits to the generalisability of the findings presented in this thesis beyond the jurisdictions and therapeutic areas examined.

Secondly, it is important to consider the contemporaneity of some of the findings presented in this thesis. In chapter 4, the sample of cancer medicines evaluated was approved from 2009-2013, with follow-up until 2018. Selecting an older sample was in line with the study aims, which included an assessment of evidence generation over the product life cycle (i.e. including a substantial follow-up period after approval to analyse the emergence of post-authorisation confirmatory studies). This study revealed acceptance of uncertainty by both EMA and FDA and analysed how this was managed by the two agencies. By contrast, chapter 5 focused on a more recent period (cancer medicine approvals from 2014-2023) and documented an increase in regulatory

acceptance of single-arm trials at the EMA. Including studies focusing on different periods therefore allows insights into patterns in regulatory decision-making that would not be observable within shorter time frames. Across chapters 4 and 5, a consistent trend of regulatory acceptance of uncertainty was observed, providing reassurance that findings from an older sample of new medicine approvals remain relevant to current policy debates.

A third limitation relates to limits to the causal interpretation of results presented in this thesis due to the methodological approaches in the three empirical chapters. However, this thesis did not set out to provide causal answers, and the methodological choices reflect the research aims of investigating the (methodological) assumptions behind regulatory approval based on non-randomised studies, and the regulatory management of the resulting uncertainty. Given the complexity of these research aims at the intersection of methodological research and regulatory policy, a mixed-methods approach was deemed appropriate. In chapter 3, a meta-epidemiological study design was used to assess potential discrepancies in effect estimates obtained from randomised vs. non-randomised studies. While the study was designed to identify closely matched pairs of studies, meta-epidemiological research is observational in nature and requires cautious interpretation of findings. For chapters 4 and 5, document analysis was used to investigate regulatory practice. Analysing data from publicly available documents, including regulatory outcomes, allows the identification of stated and revealed preferences. However, nuances in the process leading to regulatory decisions may not be reflected in these documents. Future research may address these limitations using complementary methods (see section 6.6 below).

A final limitation relates to the scope of this thesis. While it documents trends in regulatory practice and provides novel empirical insights into the validity of methodological assumptions behind regulatory acceptance of non-randomised evidence, it does not address the underlying reasons behind potential shifts in evidence standards. As introduced in the literature review (section 2.2), a range of theoretical frameworks aim to explain changes in regulatory practice, but investigating these was outside the scope of this thesis. Similarly, the thesis did not aim to directly assess the consequences of approval based on non-randomised evidence on patients and the pharmaceutical innovation system. These topics feature in the discussion section, but

the empirical evidence presented in this thesis did not directly address them. Future research is warranted to empirically analyse the net contribution of approval based on limited evidence to population health (see section 6.6 below).

To summarise, while the limitations outlined above and in the three empirical chapters restrict the interpretation of findings, they also serve to identify avenues for future research.

6.6. Directions for future research

In addressing its key research questions, this thesis has revealed some gaps that may be addressed through future research.

Qualitative research to better understand regulatory decision making

In analysing regulatory approval and the evidence base used to inform these decisions, this thesis relied on document analysis. Data was extracted from publicly available documents, including on regulatory outcomes from such as approval vs. no approval decision, the use of different regulatory pathways, final approved indications, and imposed post-marketing studies, and on acceptance of uncertain evidence, including non-randomised studies. The research was also informed by a review of regulatory guidance documents and the empirical literature on regulatory decisions to contextualise the data extracted from public documents. To complement these analyses, qualitative research may provide insights into the reasoning behind regulatory decision making. Such research, potentially including semi-structured interviews with regulators as well as participatory observation of committee meetings, could focus on the options considered by regulators when facing limited evidence, and to identify thresholds for regulatory acceptance of uncertainty. These methodological approaches are not new for understanding regulatory outcomes,[238, 318] but their application to investigate acceptance of uncertainty would provide an important contribution to the literature.

Methodological research on non-randomised studies

There is also scope for further methodological research on the internal validity of non-randomised studies. While pilot projects focusing on selected therapeutic areas with good availability of real-world data have suggested that findings of RCTs may be

replicated using non-randomised studies,[48] the meta-epidemiological study presented in chapter 3 raises important concerns about the potential for non-randomised studies to replace RCTs. The study showed that the internal validity of effect estimates obtained from the body of non-randomised studies – rather than studies for selected topics with appropriate data availability – is subject to substantial uncertainty. Due to the study design (i.e. using data from published meta-analyses), an in-depth analysis of the potential for advanced analytical methods and different data sources to reduce uncertainty was not possible, since these characteristics of non-randomised studies were not always reported in source meta-analyses. Future meta-epidemiological research could therefore focus on the impact of specific design characteristics, including analytical methods used and different types of data sources (including real-world data from disease registries, EHRs, and claims data), on the internal validity of effect estimates. Given the documented rare application of advanced analytical methods for adjusting external controls in non-randomised studies submitted for regulatory approval (see literature review in section 2.4.2), it would be particularly important to identify the existence and magnitude of discrepancies in treatment effects obtained from naïve (unadjusted) analyses vs. RCTs, as compared to the application of different methods for adjusting for potential confounding.

Future research could also investigate a potentially evolving role of non-randomised studies for regulatory approval. While the focus of previous research, including the empirical studies on the use of non-randomised studies for EMA and FDA approvals presented in chapters 4 and 5, was on evidence submitted by pharmaceutical companies and thereby only implicitly analysing regulatory standards, future research could investigate the role and methodological features of non-randomised studies designed and conducted by regulators themselves. Specifically, the EMA has conducted a pilot project on the use of real-world evidence in regulator-led studies, which includes studies conducted in a federated data network and platform for real-world data from across Europe (Data Analysis and Real World Interrogation Network / DARWIN EU).[319] These non-interventional studies can be commissioned by regulators based on questions they formulate or even conducted by regulators themselves and are intended to directly support their needs. Possible use cases include understanding disease epidemiology and the feasibility of clinical trials in a given indication, but also

addressing open questions about efficacy and safety.[319] Regulators may thus start to occupy a role in actively generating the evidence they need, including evidence to address uncertainties in the data submitted by medicines developers. Characteristics of pilot regulator-led studies so far suggest a focus on safety studies, but there is clear interest in using real-world data to estimate treatment effects.[40, 319]

It will be important to monitor the role of such studies in future approvals. Importantly, future research should evaluate possible impacts of more regulator-led studies on the cumulative evidence on new medicines generated by pharmaceutical companies across pre- and post-marketing phases. A possible dynamic that requires critical monitoring could see private companies reacting to an increased willingness of public authorities to generate the evidence required by submitting clinical trial data earlier, counting on public efforts to address remaining uncertainties. It will thus be important to analyse the characteristics of regulator-led non-interventional studies in relation to the evidence available at the time of approval, and to follow trends over time.

Population health impacts

Finally, an important topic for future research are the population health impacts of early regulatory approval based on limited evidence. These can be thought of in terms of the opportunity cost of introducing new medicines for which there is no evidence of their benefit in terms of extending or improving people's lives. Opportunity cost is a health economic concept that aims to give visibility to the fact that the benefits of some patients (e.g. through financing of a new treatment for them) may come at the cost of health foregone for other patients (e.g. through reduction of funds for other interventions).[320]

Expedited approval pathways, including those allowing approval based on less robust evidence, were introduced to address the "evidence vs. access conundrum", aiming to balance early patient access with maintaining robust evidence requirements across the product life cycle. However, empirical research conducted so far has raised questions about whether either of these two aims have been achieved. As discussed in section 6.2.3, approval based on limited evidence does not appear to have translated into earlier availability of new medicines to patients because reimbursement and pricing processes for these products take longer, or because products are not even

submitted for regular reimbursement when the evidence is considered limited. At the same time, the cumulative evidence on the therapeutic benefit of these medicines often continues to be less robust. Overall, there are therefore questions about the patient benefit of products approved based on non-randomised studies or approved through expedited pathways.

However, early approval based on limited evidence may be justified for some products which have proven to be therapeutic breakthroughs that provided patients with a meaningful new treatment option. For example, venetoclax for treatment of chronic lymphocytic leukaemia (CLL) initially received conditional approval on the basis of single-arm trial data by both EMA and FDA. In a confirmatory RCT, the product was shown to improve seven-year overall survival by 19%.[321] For many other products approved based on limited evidence, no benefits on patient-relevant outcomes have been confirmed.[155]

Despite the lack of evidence, products without confirmed benefits on overall survival or quality of life may still be used in health care, and they require resources to finance and administer them, potentially resulting in health foregone for the population as a whole. While such effects have been shown for the comparatively higher price per quality-adjusted life year (QALY) paid for medicines in the UK compared to other health interventions,[322] the population health impacts of early medicines approval based on limited evidence represent a research gap. Extending the scope of this thesis from medicines regulation to the real-world implications of availability of new medicines, future research could thus address the question: what is the net QALY gain of early regulatory approval?

7. Conclusions

The contributions of medicines regulatory bodies to protecting and improving public health are undisputed. Starting from an early focus on ex-post safety assessments in the 1960s, they have evolved and contributed significantly to evidence-based practices in medicine by ensuring that newly authorised treatments are both safe and effective, requiring robust evidence from medicines developers before their products are allowed to enter the market. In implementing these standards, regulators face a difficult trade-off between maintaining high evidence thresholds for market entry and accepting uncertainty about a new medicine's clinical benefit in order to make it available quicker to patients. Non-randomised studies may contribute to accelerated access to new medicines by shortening clinical development times before approval, and by providing confirmatory evidence on products approved based on limited, early evidence in the post-marketing phase. This contribution, however, relies on methodological assumptions about the validity of clinical evidence generated through non-randomised studies and the necessity for using non-randomised – rather than randomised – studies for approval.

This thesis investigated these assumptions and analysed how regulators manage uncertainty introduced through such studies. It showed that effect estimates obtained from RCTs and non-randomised studies frequently exhibited important discrepancies, indicating that substituting RCTs with non-randomised studies may lead to different conclusions about the benefit-risk profiles of new medicines. While some divergence between effect estimates from randomised and non-randomised studies is expected, the thesis showed that these discrepancies were frequently large enough to have implications for decision-making across a large set of clinical questions. An increased reliance on non-randomised studies for approval therefore introduces substantial uncertainty into regulatory decision making. Importantly, this uncertainty is unlikely to be fully resolved after approval due to continuing methodological weaknesses in confirmatory studies, including non-randomised study designs and, even when RCTs are imposed, the measurement of surrogate endpoints. Given the substantial uncertainty over clinical benefit for patients, the thesis raised questions whether relying on non-randomised studies for regulatory approval was justified, even for cancer

indications with small patient populations, as RCTs were shown to be feasible for most of them.

In conclusion, substituting RCTs through non-randomised studies leads to increased uncertainty about the clinical benefit of new medicines to patients, and current regulatory tools, including conditional approval, may not be able to fully resolve it. Regulators face an important challenge in managing the inherent uncertainty in clinical evidence when making approval decisions. Such uncertainty also arises in evidence generated from traditional RCTs, particularly when trials are conducted in narrow populations, measure surrogate endpoints, or use suboptimal comparators. However, when RCTs are substituted by non-randomised studies, additional uncertainties are introduced. While non-randomised studies occupy an important role in informing regulatory decisions, their methodological weaknesses in generating causal evidence on treatment effects of new medicines represent significant limits. Recognising their key role in setting evidence standards for new medicines and contributing to the availability of medicines with proven benefits, regulators should pursue efforts to make RCTs measuring patient-relevant outcomes more feasible.

References

[1] European Medicines Agency. About Us. EMA; 2025 Available from: https://www.ema.europa.eu/en/about-us/what-we-do (last access 22 July 2025)

[2] US Food and Drug Administration. What we do. FDA mission. FDA; 2023 Available from: https://www.fda.gov/aboutfda/whatwedo/#mission (last access 22 July 2025)

[3] Darrow JJ, Avorn J, Kesselheim AS. FDA Approval and Regulation of Pharmaceuticals, 1983-2018. JAMA. 2020;323(2): 164-76. Epub 2020/01/15.

[4] Greene JA, Podolsky SH. Reform, regulation, and pharmaceuticals--the Kefauver-Harris Amendments at 50. N Engl J Med. 2012;367(16): 1481-3. Epub 2012/10/19.

[5] Callréus T, Schneider CK. The Emergence of Regulatory Science in Pharmaceutical Medicine. Pharmaceutical Medicine. 2013;27(6): 345-345-51.

[6] IOM (Institute of Medicine). Building a National Framework for the Establishment of Regulatory Science for Drug Development: WorkshopSummary.; Washington, DC [online]. https://www.ncbi.nlm.nih.gov/books/NBK54404/pdf/Bookshelf_NBK54404.pdf (last access: 1 Nov 2025): The National Academies Press; 2011.

[7] IOM (Institute of Medicine). Strengthening a Workforcefor Innovative Regulatory Science in Therapeutics Development: Workshop Summary; Washington, D.C. [online]. https://www.ncbi.nlm.nih.gov/books/NBK91502/pdf/Bookshelf_NBK91502.pdf (last access: 22 Dec 2026): The National Academies Press; 2012.

[8] US Food and Drug Administration. Focus Areas of Regulatory Science - Introduction. FDA; 2022 Available from: https://www.fda.gov/science-research/focus-areas-regulatory-science-report/focus-areas-regulatory-science-introduction (last access 1 Nov 2025)

[9] Torgerson CJ, Torgerson DJ, Taylor CA. Randomized controlled trials and nonrandomized designs. In: Wholey JS, Hatry HP, Newcomer KE, Eds. 3 ed. San Francisco: Jossey-Bass; 2010.

[10] Sibbald B, Roland M. Understanding controlled trials. Why are randomised controlled trials important? BMJ. 1998;316(7126): 201. Epub 1998/02/20.

[11] Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ. 2008;336(7650): 924-6. Epub 2008/04/26.

[12] Grimes DA, Schulz KF. An overview of clinical research: The lay of the land. Lancet. 2002;359: 57-61.

[13] Higgins JPT, Savović J, Page MJ, Elbers RG, Sterne JAC. Chapter 8: Assessing risk of bias in a randomized trial. In: Higgins JPT, Thomas J, Cumpston M, Li T, Page MJ, Welch VA, Eds. Cochrane Handbook for Systematic Reviews of Interventions version 65 (updated August 2024). [online]. www.cochrane.org/handbook (last accessed 9 July 2025): Cochrane; 2024.

[14] Naci H, Davis C, Savovic J, Higgins JPT, Sterne JAC, Gyawali B, et al. Design characteristics, risk of bias, and reporting of randomised controlled trials supporting approvals of cancer drugs by European Medicines Agency, 2014-16: cross sectional analysis. BMJ. 2019;366: l5221. Epub 2019/09/20.

[15] Sterne JA, Hernan MA, Reeves BC, Savovic J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ. 2016;355: i4919. Epub 2016/10/14.

[16] European Medicines Agency. ICH Topic E 10: Choice of Control Group in Clinical Trials. London [online]. https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-10-choice-control-group-clinical-trials-step-5_en.pdf (last access 23 Sept 2025): EMA, 2001.

[17] European Medicines Agency. ICH guideline E8 (R1) on general considerations for clinical studies. EMA/CHMP/ICH/544570/1998. Amsterdam [online]. https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/ich-guideline-e8-r1-general-considerations-clinical-studies_en.pdf (last access 8 July 2025): EMA, 2021.

[18] US Food and Drug Administration. Demonstrating Substantial Evidence of Effectiveness for Human Drug and Biological Products - Guidance for Industry (Draft Guidance). Silver Spring, MD [online]. https://www.fda.gov/media/133660/download (last access 15 September 2023): FDA, 2019.

[19] US Food and Drug Administration. Clinical Trial Considerations to Support Accelerated Approval of Oncology Therapeutics. Guidance for Industry. Silver Spring, MD [online]. https://www.fda.gov/media/166431/download (last access 15 July 2025): FDA, 2023.

[20] Rosenthal R, Kasenda B, Dell-Kuster S, von Elm E, You J, Blumle A, et al. Completion and Publication Rates of Randomized Controlled Trials in Surgery: An Empirical Study. Ann Surg. 2015;262(1): 68-73. Epub 2014/07/01.

[21] Sully BG, Julious SA, Nicholl J. A reinvestigation of recruitment to randomised, controlled, multicenter trials: a review of trials funded by two UK funding agencies. Trials. 2013;14(1): 166. Epub 2013/06/14.

[22] Britton A, McKee M, Black N, McPherson K, Sanderson C, Bain C. Choosing between randomised and non-randomised studies: a systematic review. Health Technol Assess. 1998;2(13).

[23] European Medicines Agency. Reflection paper on establishing efficacy based on singlearm trials submitted as pivotal evidence in a marketing authorisation application. Considerations on evidence from single-arm trials. Amsterdam [online]. https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-establishing-efficacy-based-single-arm-trials-submitted-pivotal-evidence-marketing-authorisation-application_en.pdf (last access 23 Sept 2024): EMA, 2024.

[24] Ford I, Norrie J. Pragmatic Trials. N Engl J Med. 2016;375(5): 454-63. Epub 2016/08/16.

[25] Mathes T, Buehn S, Prengel P, Pieper D. Registry-based randomized controlled trials merged the strength of randomized controlled trails and observational studies and give rise to more pragmatic trials. J Clin Epidemiol. 2018;93: 120-7. Epub 2017/09/28.

[26] Burns L, Roux NL, Kalesnik-Orszulak R, Christian J, Hukkelhoven M, Rockhold F, et al. Real-World Evidence for Regulatory Decision-Making: Guidance From Around the World. Clin Ther. 2022;44(3): 420-37. Epub 2022/02/20.

[27] Cave A, Kurz X, Arlett P. Real-World Data for Regulatory Decision Making: Challenges and Possible Solutions for Europe. Clin Pharmacol Ther. 2019;106(1): 36-9. Epub 2019/04/11.

[28] Agrawal S, Arora S, Amiri-Kordestani L, de Claro RA, Fashoyin-Aje L, Gormley N, et al. Use of Single-Arm Trials for US Food and Drug Administration Drug Approval in Oncology, 2002-2021. JAMA Oncol. 2023;9(2): 266-72. Epub 2022/12/30.

[29] Mulder J, Teerenstra S, van Hennik PB, Pasmooij AMG, Stoyanova-Beninska V, Voest EE, et al. Single-arm trials supporting the approval of anticancer medicinal products in the European Union: contextualization of trial results and observed clinical benefit. ESMO Open. 2023;8(2): 101209. Epub 2023/04/14.

[30] Ribeiro TB, Bennett CL, Colunga-Lozano LE, Araujo APV, Hozo I, Djulbegovic B. Increasing FDA-accelerated approval of single-arm trials in oncology (1992 to 2020). J Clin Epidemiol. 2023;159: 151-8. Epub 2023/04/11.

[31] Tenhunen O, Lasch F, Schiel A, Turpeinen M. Single-Arm Clinical Trials as Pivotal Evidence for Cancer Drug Approval: A Retrospective Cohort Study of Centralized European Marketing Authorizations Between 2010 and 2019. Clin Pharmacol Ther. 2020;108(3): 653-60. Epub 2020/06/25.

[32] Franklin JM, Liaw KL, Iyasu S, Critchlow CW, Dreyer NA. Real-world evidence to support regulatory decision making: New or expanded medical product indications. Pharmacoepidemiol Drug Saf. 2021;30(6): 685-93. Epub 2021/03/07.

[33]Wallach JD, Egilman AC, Dhruva SS, McCarthy ME, Miller JE, Woloshin S, et al. Postmarket studies required by the US Food and Drug Administration for new drugs and biologics approved between 2009 and 2012: cross sectional analysis. BMJ. 2018;361: k2031. Epub 2018/05/26.

[34] Corrigan-Curay J, Martin D. Real World Evidence - From Safety to a Potential Tool for Advancing Innovative Ways to Develop New Medical Therapies. FDA; 2020 Available from: https://www.fda.gov/drugs/cder-conversations/real-world-evidence-safety-potential-tool-advancing-innovative-ways-develop-new-medical-therapies (last access 14 July 2025)

[35]European Medicines Agency. Real-world evidence framework to support EU regulatory decision-making: Report on the experience gained with regulator-led studies from September 2021 to February 2023. Amsterdam [online]. https://www.ema.europa.eu/en/documents/report/real-world-evidence-framework-support-eu-regulatory-decision-making-report-experience-gained_.pdf (last access 15 Sep 2023): European Medicines Agency, 2023.

[36] Zhang AD, Puthumana J, Downing NS, Shah ND, Krumholz HM, Ross JS. Assessment of Clinical Trials Supporting US Food and Drug Administration Approval of Novel Therapeutic Agents, 1995-2017. JAMA Netw Open. 2020;3(4): e203284. Epub 2020/04/22.

[37] Downing NS, Aminawung JA, Shah ND, Krumholz HM, Ross JS. Clinical trial evidence supporting FDA approval of novel therapeutic agents, 2005-2012. JAMA. 2014;311(4): 368-77. Epub 2014/01/23.

[38] Franklin JM, Schneeweiss S. When and How Can Real World Data Analyses Substitute for Randomized Controlled Trials? Clin Pharmacol Ther. 2017;102(6): 924-33. Epub 2017/08/25.

[39] US Food and Drug Administration. Framework for FDA's Real-World Evidence Program. Silver Spring, MD [online]. https://www.fda.gov/media/120060/download (last access 3 Nov 2022): FDA, 2018.

[40] European Medicines Agency. Real-world evidence framework to support EU regulatory decision-making: 3rd report on the experience gained with regulator-led studies from February 2024 to February 2025. Amsterdam [online]. https://www.ema.europa.eu/en/documents/report/real-world-evidence-framework-support-eu-regulatory-decision-making-3rd-report-experience-gained-regulator-led-studies-february-2024-february-2025_en.pdf (last access 11 July 2025): EMA, 2025.

[41] European Medicines Agency. Real-world evidence provided by EMA - Support for regulatory decision-making. Amsterdam [online]. https://www.ema.europa.eu/en/documents/other/guide-real-world-evidence-provided-ema-support-regulatory-decision-making_en.pdf (last access 11 July 2025): EMA, 2024.

[42] Hill AB. The environment and disease: association or causation? 1965. J R Soc Med. 2015;108(1): 32-7. Epub 2015/01/13.

[43] Holland PW. Statistics and Causal Inference. Journal of the American Statistical Association. 1986;81(396): 945-60.

[44] Agoritsas T, Merglen A, Shah ND, O'Donnell M, Guyatt GH. Adjusted Analyses in Studies Addressing Therapy and Harm: Users' Guides to the Medical Literature. JAMA. 2017;317(7): 748-59.

[45] Kreif N, Grieve R, Hangartner D, Turner AJ, Nikolova S, Sutton M. Examination of the Synthetic Control Method for Evaluating Health Policies with Multiple Treated Units. Health Econ. 2016;25(12): 1514-28. Epub 2015/10/08.

[46] Kreif N, Grieve R, Sadique MZ. Statistical methods for cost-effectiveness analyses that use observational data: a critical appraisal tool and review of current practice. Health Econ. 2013;22(4): 486-500. Epub 2012/03/27.

[47] Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behav Res. 2011;46(3): 399-424. Epub 2011/08/06.

[48] Wang SV, Schneeweiss S, Initiative R-D, Franklin JM, Desai RJ, Feldman W, et al. Emulation of Randomized Clinical Trials With Nonrandomized Database Analyses: Results of 32 Clinical Trials. JAMA. 2023;329(16): 1376-85. Epub 2023/04/25.

[49] Hernan MA, Wang W, Leaf DE. Target Trial Emulation: A Framework for Causal Inference From Observational Data. JAMA. 2022;328(24): 2446-7. Epub 2022/12/13.

[50] Etheredge LM. A rapid-learning health system. Health Aff (Millwood). 2007;26(2): w107-18. Epub 2007/01/30.

[51] Eichler HG, Bloechl-Daum B, Broich K, Kyrle PA, Oderkirk J, Rasi G, et al. Data Rich, Information Poor: Can We Use Electronic Health Records to Create a Learning Healthcare System for Pharmaceuticals? Clin Pharmacol Ther. 2019;105(4): 912-22. Epub 2018/09/05.

[52] European Medicines Agency. Guideline on registry-based studies. Amsterdam [online]. https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-registry-based-studies_en-0.pdf (last accessed 3 Nov 2022): EMA, 2021.

[53]McGettigan P, Alonso Olmo C, Plueschke K, Castillon M, Nogueras Zondag D, Bahri P, et al. Patient Registries: An Underused Resource for Medicines Evaluation : Operational proposals for increasing the use of patient registries in regulatory assessments. Drug Saf. 2019;42(11): 1343-51. Epub 2019/07/16.

[54] Feinberg BA, Gajra A, Zettler ME, Phillips TD, Phillips EG, Jr., Kish JK. Use of Real-World Evidence to Support FDA Approval of Oncology Drugs. Value Health. 2020;23(10): 1358-65. Epub 2020/10/10.

[55] Kesselheim AS, Wang B, Franklin JM, Darrow JJ. Trends in utilization of FDA expedited drug development and approval programs, 1987-2014: cohort study. BMJ. 2015;351: h4633. Epub 2015/09/25.

[56] Kimmit R, Vieira M. Knowledge Portal on Innovation and Access to Medicines, Research Synthesis: Time and success rate of pharmaceutical R&D. Geneva [online]. https://www.knowledgeportalia.org/r-d-time-and-success-rate (last access 10 July 2025): Graduate Institute, 2020.

[57] European Medicines Agency. EMA Regulatory Science to 2025: Stategic Reflection. London [online]. https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/ema-regulatory-science-2025-strategic-reflection_en.pdf (last accessed 10 July 2025): EMA, 2018.

[58] US Food and Drug Administration. Advancing Regulatory Science at FDA: A Strategic Plan. Silver Spring, MD [online]. https://www.fda.gov/media/81109/download (last access 15 Nov 2022): FDA, 2011.

[59] Eisenberg RS. The Role of the FDA in Innovation Policy. Michigan Telecommunications and Technology Law Review. 2007;13(2): 345-88.

[60] Kapczynski A. Dangerous Times: The FDA's Role in Information Production, Past and Future. Minnesota Law Review. 2018;102: 2357-82.

[61]Naci H, Smalley KR, Kesselheim AS. Characteristics of Preapproval and Postapproval Studies for Drugs Granted Accelerated Approval by the US Food and Drug Administration. JAMA. 2017;318(7): 626-36. Epub 2017/08/16.

[62] Naci H, Wouters OJ, Gupta R, Ioannidis JPA. Timing and Characteristics of Cumulative Evidence Available on Novel Therapeutic Agents Receiving Food and Drug Administration Accelerated Approval. Milbank Q. 2017;95(2): 261-90. Epub 2017/06/08.

[63] Eichler HG, Baird LG, Barker R, Bloechl-Daum B, Borlum-Kristensen F, Brown J, et al. From adaptive licensing to adaptive pathways: delivering a flexible life-span approach to bring new drugs to patients. Clin Pharmacol Ther. 2015;97(3): 234-46. Epub 2015/02/12.

[64] Eichler HG, Pignatti F, Flamion B, Leufkens H, Breckenridge A. Balancing early market access to new drugs with the need for benefit/risk data: a mounting dilemma. Nat Rev Drug Discov. 2008;7(10): 818-26. Epub 2008/09/13.

[65] Herder M. Pharmaceutical Drugs of Uncertain Value, Lifecycle Regulation at the US Food and Drug Administration, and Institutional Incumbency. Milbank Q. 2019;97(3): 820-57. Epub 2019/08/14.

[66] Baird LG, Banken R, Eichler HG, Kristensen FB, Lee DK, Lim JC, et al. Accelerated access to innovative medicines for patients in need. Clin Pharmacol Ther. 2014;96(5): 559-71. Epub 2014/07/10.

[67] Franco P, Jain R, Rosenkrands-Lange E, Hey C, Koban MU. Regulatory Pathways Supporting Expedited Drug Development and Approval in ICH Member Countries. Ther Innov Regul Sci. 2023;57(3): 484-514. Epub 2022/12/04.

[68] DiMasi JA, Faden L. Factors Associated with Multiple FDA Review Cycles and Approval Phase Times. Drug Information Journal. 2009;43(2): 201-25.

[69] Downing NS, Aminawung JA, Shah ND, Braunstein JB, Krumholz HM, Ross JS. Regulatory review of novel therapeutics--comparison of three regulatory agencies. N Engl J Med. 2012;366(24): 2284-93. Epub 2012/05/18.

[70] Banzi R, Gerardi C, Bertele V, Garattini S. Approvals of drugs with uncertain benefit-risk profiles in Europe. Eur J Intern Med. 2015;26(8): 572-84. Epub 2015/09/08.

[71] Davis C, Lexchin J, Jefferson T, Gotzsche P, McKee M. "Adaptive pathways" to drug authorisation: adapting to industry? BMJ. 2016;354: i4437. Epub 2016/08/18.

[72] Ermisch M, Bucsics A, Vella Bonanno P, Arickx F, Bybau A, Bochenek T, et al. Payers' Views of the Changes Arising through the Possible Adoption of Adaptive Pathways. Front Pharmacol. 2016;7: 305. Epub 2016/10/14.

[73] Kesselheim AS, Avorn J. New "21st Century Cures" Legislation: Speed and Ease vs Science. JAMA. 2017;317(6): 581-2. Epub 2017/01/06.

[74] Hoekman J, Klamer TT, Mantel-Teeuwisse AK, Leufkens HG, De Bruin ML. Characteristics and follow-up of postmarketing studies of conditionally authorized medicines in the EU. Br J Clin Pharmacol. 2016;82(1): 213-26. Epub 2016/03/19.

[75] Kepplinger EE. FDA's Expedited Approval Mechanisms for New Drug Products. Biotechnol Law Rep. 2015;34(1): 15-37. Epub 2015/02/26.

[76] Wallach JD, Ross JS, Naci H. The US Food and Drug Administration's expedited approval programs: Evidentiary standards, regulatory trade-offs, and potential improvements. Clin Trials. 2018;15(3): 219-29. Epub 2018/06/07.

[77] Darrow JJ, Avorn J, Kesselheim AS. The FDA Breakthrough-Drug Designation — Four Years of Experience. New England Journal of Medicine. 2018;378(15): 1444-53.

[78] Salcher-Konrad M, Naci H. Unintended Consequences of Coverage Laws Targeting Cancer Drugs. J Law Med Ethics. 2020;48(3): 552-4. Epub 2020/10/07.

[79] Leopold C, Haffajee RL, Lu CY, Wagner AK. The Complex Cancer Care Coverage Environment - What is the Role of Legislation? A Case Study from Massachusetts. J Law Med Ethics. 2020;48(3): 538-51. Epub 2020/10/07.

[80] Xander NSH, Belleman T, Salcher-Konrad M, Hendrickx A, Chen J, Klein Gebbink AS, et al. Price determinants and pricing policies concerning potentially innovative health technologies: a scoping review. Eur J Health Econ. 2025. Epub 2025/09/06.

[81] Kemp R, Prasad V. Surrogate endpoints in oncology: when are they acceptable for regulatory and clinical decisions, and are they currently overused? BMC Medicine. 2017;15(1): 134.

[82] Beaver JA, Howie LJ, Pelosof L, Kim T, Liu J, Goldberg KB, et al. A 25-Year Experience of US Food and Drug Administration Accelerated Approval of Malignant Hematology and Oncology Drugs and Biologics: A Review. JAMA Oncol. 2018;4(6): 849-56. Epub 2018/03/02.

[83] European Medicines Agency. Enhertu EPAR CHMP assessment report EMA/2446/2021. Amsterdam: EMA, 2020.

[84] Andre F, Hee Park Y, Kim SB, Takano T, Im SA, Borges G, et al. Trastuzumab deruxtecan versus treatment of physician's choice in patients with HER2-positive metastatic breast cancer (DESTINY-Breast02): a randomised, open-label, multicentre, phase 3 trial. Lancet. 2023;401(10390): 1773-85. Epub 2023/04/23.

[85] Davis C, Naci H, Gurpinar E, Poplavska E, Pinto A, Aggarwal A. Availability of evidence of benefits on overall survival and quality of life of cancer drugs approved by European Medicines Agency: retrospective cohort study of drug approvals 2009-13. BMJ. 2017;359: j4530. Epub 2017/10/06.

[86] Timmermans K. Monopolizing clinical trial data: implications and trends. PLoS Med. 2007;4(2): e2. Epub 2007/02/15.

[87] Hickey KJ, Ward EH. The Role of Patents and Regulatory Exclusivities in Drug Pricing. [online]. https://www.congress.gov/crs_external_products/R/PDF/R46679/R46679.6.pdf (last accessed 8 July 2025): Congressional Research Service, 2024.

[88] European Medicines Agency. Adcetris EPAR assessment report EMEA/H/C/002455. London: EMA, 2012.

[89] F. D. A. Center for Drug Evaluation Research. Cross discipline team leader review: Adcetris/brentuximab vedotin BLA 125388. 2011.

[90] European Medicines Agency. Lytgobi EPAR assessment report EMA/218116/2023. Amsterdam: EMA, 2023.

[91] European Medicines Agency. Retsevmo EPAR assessment report EMA/9037/2021. Amsterdam: EMA, 2021.

[92] European Medicines Agency. Elzonris EPAR assessment report EMA/CHMP/24699/2021. Amsterdam: EMA, 2020.

[93] Grieve R, Abrams K, Claxton K, Goldacre B, James N, Nicholl J, et al. Cancer Drugs Fund requires further reform. BMJ (Clinical research ed). 2016;354: i5090.

[94] Ioannidis JP. Contradicted and initially stronger effects in highly cited clinical research. JAMA. 2005;294(2): 218-28. Epub 2005/07/15.

[95] Ioannidis JP, Haidich AB, Pappa M, Pantazis N, Kokori SI, Tektonidou MG, et al. Comparison of evidence of treatment effects in randomized and nonrandomized studies. JAMA. 2001;286(7): 821-30. Epub 2001/08/22.

[96] Hemkens LG, Contopoulos-Ioannidis DG, Ioannidis JP. Agreement of treatment effects for mortality from routinely collected data and subsequent randomized trials: meta-epidemiological survey. BMJ. 2016;352: i493. Epub 2016/02/10.

[97] Moore TJ, Furberg CD. Electronic Health Data for Postmarket Surveillance: A Vision Not Realized. Drug Saf. 2015;38(7): 601-10. Epub 2015/05/31.

[98] US Food and Drug Administration. Demonstrating Substantial Evidence of Effectiveness With One Adequate and Well-Controlled Clinical Investigation and Confirmatory Evidence. Guidance for Industry. Silver Spring, MD [online]. https://www.fda.gov/media/172166/download (last access 27 Sep 2023): FDA, 2023.

[99] Mulinari S, Davis C. The will of Congress? Permissive regulation and the strategic use of labeling for the anti-influenza drug Relenza. Soc Stud Sci. 2020;50(1): 145-69. Epub 2019/11/20.

[100] Davis C, Abraham J. Desperately seeking cancer drugs: explaining the emergence and outcomes of accelerated pharmaceutical regulation. Sociol Health Illn. 2011;33(5): 731-47. Epub 2011/02/15.

[101] Davis C, Abraham J. Putting Pharmaceutical Regulation to the Test: From Historical Description to a Social Science for Public Health. Palgrave Macmillan, London; 2013. p. 5-35.

[102] Borup R, Traulsen JM, Kaae S. Regulatory Capture in Pharmaceutical Policy Making: The Case of National Medicine Agencies Related to the EU Falsified Medicines Directive. Pharmaceut Med. 2019;33(3): 199-207. Epub 2020/01/15.

[103] Carpenter D. Reputation and Power: Organizational Image and Pharmaceutical Regulation at the FDA; Princeton: Princeton University Press; 2010.

[104] Carpenter DP. Protection without Capture: Product Approval by a Politically Responsive, Learning Regulator. American Political Science Review. 2004;98(4): 613-31.

[105] Davis C, Abraham J. Unhealthy Pharmaceutical Regulation: Innovation, Politics and Promissory Science; Basingstoke: Palgrave Macmillan; 2013.

[106] Fisher JA. Coming Soon to a Physician Near You: Medical Neoliberalism and Pharmaceutical Clinical Trials. Harvard Health Policy Rev. 2007;8(1): 61-70. Epub 2007/04/01.

[107] Sterne JA, Juni P, Schulz KF, Altman DG, Bartlett C, Egger M. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. Stat Med. 2002;21(11): 1513-24. Epub 2002/07/12.

[108] Savovic J, Jones H, Altman D, Harris R, Juni P, Pildal J, et al. Influence of reported study design characteristics on intervention effect estimates from randomised controlled trials: combined analysis of meta-epidemiological studies. Health Technol Assess. 2012;16(35): 1-82. Epub 2012/09/20.

[109] Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, et al. Evaluating non-randomised intervention studies. Health Technol Assess. 2003;7(27): iii-x, 1-173. Epub 2003/09/23.

[110] Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. New England Journal of Medicine. 2000;342: 1878-86.

[111] Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. N Engl J Med. 2000;342(25): 1887-92. Epub 2000/06/22.

[112] Toews I, Anglemyer A, Nyirenda JL, Alsaid D, Balduzzi S, Grummich K, et al. Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials: a meta-epidemiological study. Cochrane Database Syst Rev. 2024;1(1): MR000034. Epub 2024/01/04.

[113] Edwards JP, Kelly EJ, Lin Y, Lenders T, Ghali WA, Graham AJ. Meta-analytic comparison of randomized and nonrandomized studies of breast cancer surgery. Can J Surg. 2012;55(3): 155-62. Epub 2012/03/28.

[114] Mueller D, Sauerland S, Neugebauer EA, Immenroth M. Reported effects in randomized controlled trials were compared with those of nonrandomized trials in cholecystectomy. J Clin Epidemiol. 2010;63(10): 1082-90. Epub 2010/03/30.

[115] Shikata S, Nakayama T, Noguchi Y, Taji Y, Yamagishi H. Comparison of effects in randomized controlled trials with observational studies in digestive surgery. Ann Surg. 2006;244(5): 668-76. Epub 2006/10/25.

[116] Frangakis C. The calibration of treatment effects from clinical trials to target populations. Clin Trials. 2009;6(2): 136-40. Epub 2009/04/04.

[117] Hordijk-Trion M, Lenzen M, Wijns W, de Jaegere P, Simoons ML, Scholte op Reimer WJ, et al. Patients enrolled in coronary intervention trials are not representative of patients in clinical practice: results from the Euro Heart Survey on Coronary Revascularization. Eur Heart J. 2006;27(6): 671-8. Epub 2006/01/21.

[118] Bolislis WR, Fay M, Kuhler TC. Use of Real-world Data for New Drug Applications and Line Extensions. Clin Ther. 2020;42(5): 926-38. Epub 2020/04/29.

[119] Mofid S, Bolislis WR, Kuhler TC. Real-World Data in the Postapproval Setting as Applied by the EMA and the US FDA. Clin Ther. 2022;44(2): 306-22. Epub 2022/01/26.

[120] Patel D, Grimson F, Mihaylova E, Wagner P, Warren J, van Engen A, et al. Use of External Comparators for Health Technology Assessment Submissions Based on Single-Arm Trials. Value Health. 2021;24(8): 1118-25. Epub 2021/08/11.

[121] Purpura CA, Garry EM, Honig N, Case A, Rassen JA. The Role of Real-World Evidence in FDA-Approved New Drug and Biologics License Applications. Clin Pharmacol Ther. 2022;111(1): 135-44. Epub 2021/11/03.

[122] Wang X, Dormont F, Lorenzato C, Latouche A, Hernandez R, Rouzier R. Current perspectives for external control arms in oncology clinical trials: Analysis of EMA approvals 2016-2021. J Cancer Policy. 2023;35: 100403. Epub 2023/01/17.

[123] Zettler ME. The use of real-world evidence to support FDA post-approval study requirements for oncology drugs. Expert Rev Anticancer Ther. 2022;22(6): 657-66. Epub 2022/05/06.

[124] Makady A, Ham RT, de Boer A, Hillege H, Klungel O, Goettsch W, et al. Policies for Use of Real-World Data in Health Technology Assessment (HTA): A Comparative Study of Six HTA Agencies. Value Health. 2017;20(4): 520-32. Epub 2017/04/15.

[125] RCT DUPLICATE. RCT DUPLICATE - Home. Available from: https://www.rctduplicate.org/ (last access 23 July 2025)

[126] Hansford HJ, Cashin AG, Jones MD, Swanson SA, Islam N, Douglas SRG, et al. Reporting of Observational Studies Explicitly Aiming to Emulate Randomized Trials: A Systematic Review. JAMA Netw Open. 2023;6(9): e2336023. Epub 2023/09/27.

[127] Simon-Tillaux N, Martin GL, Hajage D, Scheifer C, Beydon M, Dechartres A, et al. Conducting observational analyses with the target trial emulation approach: a methodological systematic review. BMJ Open. 2024;14(11): e086595. Epub 2024/11/13.

[128] Janda GS, Wallach JD, Dhodapkar MM, Ramachandran R, Ross JS. Feasibility of Emulating Clinical Trials Supporting US FDA Supplemental Indication Approvals of Drugs and Biologics. JAMA Intern Med. 2023;183(11): 1271-3. Epub 2023/10/02.

[129] European Medicines Agency. Reflection paper on use of real-world data in noninterventional studies to generate real-world evidence for regulatory purposes. Amsterdam [online]. https://www.ema.europa.eu/en/documents/other/reflection-paper-use-real-world-data-non-interventional-studies-generate-real-world-evidence-regulatory-purposes_en.pdf (last access 15 July 2025): EMA, 2025.

[130] US Food and Drug Administration. Real-World Evidence: Considerations Regarding Non-Interventional Studies for Drug and Biological Products. Guidance for Industry. Draft guidance. Silver Springs, MD [online]. https://www.fda.gov/media/177128/download (last access 15 July 2025): FDA, 2024.

[131] Guyatt GH, Oxman AD, Sultan S, Glasziou P, Akl EA, Alonso-Coello P, et al. GRADE guidelines: 9. Rating up the quality of evidence. J Clin Epidemiol. 2011;64(12): 1311-6. Epub 2011/08/02.

[132] Arondekar B, Duh MS, Bhak RH, DerSarkissian M, Huynh L, Wang K, et al. Real-World Evidence in Support of Oncology Product Registration: A Systematic Review of New Drug Application and Biologics License Application Approvals from 2015-2020. Clin Cancer Res. 2022;28(1): 27-35. Epub 2021/10/21.

[133] Goring S, Taylor A, Muller K, Li TJJ, Korol EE, Levy AR, et al. Characteristics of non-randomised studies using comparisons with external controls submitted for regulatory approval in the USA and Europe: a systematic review. BMJ Open. 2019;9(2): e024895. Epub 2019/03/02.

[134] Sola-Morales O, Curtis LH, Heidt J, Walsh L, Casso D, Oliveria S, et al. Effectively Leveraging RWD for External Controls: A Systematic Literature Review of

Regulatory and HTA Decisions. Clin Pharmacol Ther. 2023;114(2): 325-55. Epub 2023/04/20.

[135] Oxnard GR, Wilcox KH, Gonen M, Polotsky M, Hirsch BR, Schwartz LH. Response Rate as a Regulatory End Point in Single-Arm Studies of Advanced Solid Tumors. JAMA Oncol. 2016;2(6): 772-9. Epub 2016/02/26.

[136] Djulbegovic B, Glasziou P, Klocksieben FA, Reljic T, VanDenBergh M, Mhaskar R, et al. Larger effect sizes in nonrandomized studies are associated with higher rates of EMA licensing approval. J Clin Epidemiol. 2018;98: 24-32. Epub 2018/02/13.

[137] Razavi M, Glasziou P, Klocksieben FA, Ioannidis JPA, Chalmers I, Djulbegovic B. US Food and Drug Administration Approvals of Drugs and Devices Based on Nonrandomized Clinical Trials: A Systematic Review and Meta-analysis. JAMA Netw Open. 2019;2(9): e1911111. Epub 2019/09/12.

[138] Kersting C, Kneer M, Barzel A. Patient-relevant outcomes: what are we talking about? A scoping review to improve conceptual clarity. BMC Health Serv Res. 2020;20(1): 596. Epub 2020/07/01.

[139] Chen EY, Joshi SK, Tran A, Prasad V. Estimation of Study Time Reduction Using Surrogate End Points Rather Than Overall Survival in Oncology Clinical Trials. JAMA Intern Med. 2019;179(5): 642-7. Epub 2019/04/02.

[140] Ciani O, Buyse M, Drummond M, Rasi G, Saad ED, Taylor RS. Time to Review the Role of Surrogate End Points in Health Policy: State of the Art and the Way Forward. Value Health. 2017;20(3): 487-95. Epub 2017/03/16.

[141] Ciani O, Buyse M, Garside R, Pavey T, Stein K, Sterne JA, et al. Comparison of treatment effect sizes associated with surrogate and final patient relevant outcomes in randomised controlled trials: meta-epidemiological study. BMJ. 2013;346(7898): f457. Epub 2013/01/31.

[142] Chen EY, Raghunathan V, Prasad V. An Overview of Cancer Drugs Approved by the US Food and Drug Administration Based on the Surrogate End Point of Response Rate. JAMA Intern Med. 2019;179(7): 915-21. Epub 2019/05/29.

[143] Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? Ann Intern Med. 1996;125(7): 605-13. Epub 1996/10/01.

[144] Schuster Bruce C, Brhlikova P, Heath J, McGettigan P. The use of validated and nonvalidated surrogate endpoints in two European Medicines Agency expedited approval pathways: A cross-sectional study of products authorised 2011-2018. PLoS Med. 2019;16(9): e1002873. Epub 2019/09/11.

[145] Haslam A, Hey SP, Gill J, Prasad V. A systematic review of trial-level meta-analyses measuring the strength of association between surrogate end-points and overall survival in oncology. Eur J Cancer. 2019;106: 196-211. Epub 2018/12/12.

[146] Prasad V, Kim C, Burotto M, Vandross A. The Strength of Association Between Surrogate End Points and Survival in Oncology: A Systematic Review of Trial-Level Meta-analyses. JAMA Intern Med. 2015;175(8): 1389-98. Epub 2015/06/23.

[147] Walia A, Haslam A, Prasad V. FDA validation of surrogate endpoints in oncology: 2005-2022. J Cancer Policy. 2022;34: 100364. Epub 2022/09/27.

[148] Manyara AM, Davies P, Stewart D, Weir CJ, Young AE, Wells V, et al. Definitions, acceptability, limitations, and guidance in the use and reporting of surrogate end points in trials: a scoping review. J Clin Epidemiol. 2023;160: 83-99. Epub 2023/06/29.

[149] Kesselheim AS, Myers JA, Avorn J. Characteristics of clinical trials to support approval of orphan vs nonorphan drugs for cancer. JAMA. 2011;305(22): 2320-6. Epub 2011/06/07.

[150] Chen EY, Haslam A, Prasad V. FDA Acceptance of Surrogate End Points for Cancer Drug Approval: 1992-2019. JAMA Intern Med. 2020;180(6): 912-4. Epub 2020/04/14.

[151] Kordecka A, Walkiewicz-Zarek E, Lapa J, Sadowska E, Kordecki M. Selection of Endpoints in Clinical Trials: Trends in European Marketing Authorization Practice in Oncological Indications. Value Health. 2019;22(8): 884-90. Epub 2019/08/21.

[152] Kim C, Prasad V. Strength of Validation for Surrogate End Points Used in the US Food and Drug Administration's Approval of Oncology Drugs. Mayo Clin Proc. 2016;91(6): 713-25. Epub 2016/05/30.

[153] Wallach JD, Ciani O, Pease AM, Gonsalves GS, Krumholz HM, Taylor RS, et al. Comparison of treatment effect sizes from pivotal and postapproval trials of novel therapeutics approved by the FDA based on surrogate markers of disease: a meta-epidemiological study. BMC Med. 2018;16(1): 45. Epub 2018/03/23.

[154] Liu ITT, Kesselheim AS, Cliff ERS. Clinical Benefit and Regulatory Outcomes of Cancer Drugs Receiving Accelerated Approval. JAMA. 2024;331(17): 1471-9. Epub 2024/04/07.

[155] Gyawali B, Hey SP, Kesselheim AS. Assessment of the Clinical Benefit of Cancer Drugs Receiving Accelerated Approval. JAMA Intern Med. 2019;179(7): 906-13. Epub 2019/05/29.

[156] European Medicines Agency. Conditional marketing authorisation: Report on ten years of experience at the European Medicines Agency. London [online]. https://www.ema.europa.eu/en/documents/report/conditional-marketing-authorisation-report-ten-years-experience-european-medicines-agency_en.pdf (last access 25 July 2025): EMA, 2017.

[157] Hatswell AJ, Baio G, Berlin JA, Irs A, Freemantle N. Regulatory approval of pharmaceuticals without a randomised controlled study: analysis of EMA and FDA approvals 1999-2014. BMJ Open. 2016;6(6): e011666. Epub 2016/07/02.

[158] Ribeiro TB, Colunga-Lozano LE, Araujo APV, Bennett CL, Hozo I, Djulbegovic B. Single-arm clinical trials that supported FDA accelerated approvals have modest effect sizes and were at high risk of bias. J Clin Epidemiol. 2022;148: 193-5. Epub 2022/01/31.

[159] Rittberg R, Czaykowski P, Niraula S. Feasibility of Randomized Controlled Trials for Cancer Drugs Approved by the Food and Drug Administration Based on Single-Arm Studies. JNCI Cancer Spectr. 2021;5(4). Epub 2021/08/20.

[160] Gaddipati H, Liu K, Pariser A, Pazdur R. Rare cancer trial design: lessons from FDA approvals. Clin Cancer Res. 2012;18(19): 5172-8. Epub 2012/06/22.

[161] Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. [MB13-01] Bewertung und Auswertung von Studien bei seltenen Erkrankungen - Rapid Report. Cologne [online]. https://www.iqwig.de/download/MB13-01_Rapid-Report_Studien-bei-seltenen-Erkrankungen.pdf (last access 6 Nov 2024): IQWiG, 2014.

[162] US Food and Drug Administration. Guidance for Industry: E 10 Choice of Control Group and Related Issues in Clinical Trials. Rockville, MD [online]. https://www.fda.gov/media/71349/download (last access 15 July 2025): FDA, 2001.

[163] Stuart EA. Matching methods for causal inference: A review and a look forward. Stat Sci. 2010;25(1): 1-21. Epub 2010/09/28.

[164] Stukel TA, Fisher ES, Wennberg DE, Alter DA, Gottlieb DJ, Vermeulen MJ. Analysis of observational studies in the presence of treatment selection bias: effects of invasive cardiac management on AMI survival using propensity score and instrumental variable methods. JAMA. 2007;297(3): 278-85. Epub 2007/01/18.

[165] Abrahami D, Pradhan R, Yin H, Honig P, Baumfeld Andre E, Azoulay L. Use of Real-World Data to Emulate a Clinical Trial and Support Regulatory Decision Making: Assessing the Impact of Temporality, Comparator Choice, and Method of Adjustment. Clin Pharmacol Ther. 2021;109(2): 452-61. Epub 2020/08/09.

[166] Bosco JL, Silliman RA, Thwin SS, Geiger AM, Buist DS, Prout MN, et al. A most stubborn bias: no adjustment method fully resolves confounding by indication in observational studies. J Clin Epidemiol. 2010;63(1): 64-74. Epub 2009/05/22.

[167] Hatswell AJ, Freemantle N, Baio G. Economic Evaluations of Pharmaceuticals Granted a Marketing Authorisation Without the Results of Randomised Trials: A Systematic Review and Taxonomy. PharmacoEconomics. 2017;35(2): 163-76. Epub 2016/10/26.

[168] Kent S, Salcher-Konrad M, Boccia S, Bouvy JC, Waure C, Espin J, et al. The use of nonrandomized evidence to estimate treatment effects in health technology assessment. J Comp Eff Res. 2021;10(14): 1035-43. Epub 2021/07/20.

[169] von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Prev Med. 2007;45(4): 247-51. Epub 2007/10/24.

[170] Brown JP, Wing K, Evans SJ, Bhaskaran K, Smeeth L, Douglas IJ. Use of real-world evidence in postmarketing medicines regulation in the European Union: a systematic assessment of European Medicines Agency referrals 2013-2017. BMJ Open. 2019;9(10): e028133. Epub 2019/10/31.

[171] European Medicines Agency. Journey towards a roadmap for regulatory guidance on real-world evidence. Amsterdam [online]. https://www.ema.europa.eu/en/documents/other/journey-towards-roadmap-regulatory-guidance-real-world-evidence_en.pdf (last access 15 July 2025): EMA, 2025.

[172] US Food and Drug Administration. Considerations for the Use of Real-World Data and Real-World Evidence to Support Regulatory Decision-Making for Drug and Biological Products. Guidance for Industry. Silver Spring, MD [online]. https://www.fda.gov/media/171667/download (last access 15 July 2025): FDA, 2023.

[173] European Medicines Agency. Data Quality Framework for EU medicines regulation: application to Real-World Data. Draft. Amsterdam [online]. https://www.ema.europa.eu/en/documents/other/draft-data-quality-framework-eu-medicines-regulation-application-real-world-data_en.pdf (last access 15 July 2025): EMA, 2024.

[174] US Food and Drug Administration. Data Standards for Drug and Biological Product Submissions Containing Real-World Data. Guidance for Industry. Silver Springs, MD [online]. https://www.fda.gov/media/153341/download (last access 15 July 2025): FDA, 2023.

[175] Flynn R, Plueschke K, Quinten C, Strassmann V, Duijnhoven RG, Gordillo-Maranon M, et al. Marketing Authorization Applications Made to the European Medicines Agency in 2018-2019: What was the Contribution of Real-World Evidence? Clin Pharmacol Ther. 2022;111(1): 90-7. Epub 2021/10/25.

[176] Eskola SM, Leufkens HGM, Bate A, De Bruin ML, Gardarsdottir H. Use of Real-World Data and Evidence in Drug Development of Medicinal Products Centrally Authorized in Europe in 2018-2019. Clin Pharmacol Ther. 2022;111(1): 310-20. Epub 2021/10/25.

[177] European Medicines Agency. Guideline on the clinical evaluation of anticancer medicinal products - revision 6. Amsterdam [online]. https://www.ema.europa.eu/en/documents/scientific-guideline/draft-guideline-evaluation-anticancer-medicinal-products-man-revision-6_en.pdf (last access 31 July 2025): EMA, 2019.

[178] Cox EM, Edmund AV, Kratz E, Lockwood SH, Shankar A. Regulatory Affairs 101: Introduction to Expedited Regulatory Pathways. Clin Transl Sci. 2020;13(3): 451-61. Epub 2020/01/08.

[179] Hoekman J, Boon W. Changing standards for drug approval: A longitudinal analysis of conditional marketing authorisation in the European Union. Soc Sci Med. 2019;222: 76-83. Epub 2019/01/04.

[180] Ribeiro TB, Buss L, Wayant C, Nobre MRC. Comparison of FDA accelerated vs regular pathway approvals for lung cancer treatments between 2006 and 2018. PLoS One. 2020;15(7): e0236345. Epub 2020/07/25.

[181] Alipour-Haris G, Liu X, Acha V, Winterstein AG, Burcu M. Real-world evidence to support regulatory submissions: A landscape review and assessment of use cases. Clin Transl Sci. 2024;17(8): e13903. Epub 2024/08/02.

[182] Kashoki M, Hanaizi Z, Yordanova S, Vesely R, Bouygues C, Llinares J, et al. A Comparison of EMA and FDA Decisions for New Drug Marketing Applications 2014-2016: Concordance, Discordance, and Why. Clin Pharmacol Ther. 2020;107(1): 195-202. Epub 2019/07/16.

[183] Xie J, Li J, Liu Y, Wang H, Wang Y, Yang Y, et al. Comparison of novel oncology drugs that received dual approval from the US accelerated approval and EU conditional marketing authorisation pathways, 2006-2021: a cross-sectional study. BMJ Open. 2023;13(6): e069132. Epub 2023/06/08.

[184] Kurian M, Ferrell WJ, Ulloa Perez E, Hubbard R, Joffe S, Mamtani R, et al. Specificity of Food and Drug Administration postmarketing requirements and

associations with timely submissions and regulatory decisions for oncology accelerated approvals, 2011-2023: a cross-sectional analysis. BMJ Oncol. 2025;4(1): e000659. Epub 2025/06/16.

[185] Cherla A, Mossialos E, Salcher-Konrad M, Kesselheim AS, Naci H. Post-Marketing Requirements for Cancer Drugs Approved by the European Medicines Agency, 2004-2014. Clin Pharmacol Ther. 2022;112(4): 846-52. Epub 2022/06/07.

[186] Zeitoun JD, Ross JS, Atal I, Vivot A, Downing NS, Baron G, et al. Postmarketing studies for novel drugs approved by both the FDA and EMA between 2005 and 2010: a cross-sectional study. BMJ Open. 2017;7(12): e018587. Epub 2017/12/24.

[187] Thomson S, Witzke N, Gyawali B, Delos Santos S, Udayakumar S, Cardone C, et al. Assessing the benefit of cancer drugs approved by the European Medicines Agency using the European Society for Medical Oncology Magnitude of Clinical Benefit Scale over time. Eur J Cancer. 2021;150: 203-10. Epub 2021/05/02.

[188] Grossmann N, Wolf S, Rothschedl E, Wild C. Twelve years of European cancer drug approval-a systematic investigation of the 'magnitude of clinical benefit'. ESMO Open. 2021;6(3): 100166. Epub 2021/06/05.

[189] Hwang TJ, Ross JS, Vokinger KN, Kesselheim AS. Association between FDA and EMA expedited approval programs and therapeutic value of new medicines: retrospective cohort study. BMJ. 2020;371: m3434. Epub 2020/10/09.

[190] Vokinger KN, Kesselheim AS, Glaus CEG, Hwang TJ. Therapeutic Value of Drugs Granted Accelerated Approval or Conditional Marketing Authorization in the US and Europe From 2007 to 2021. JAMA Health Forum. 2022;3(8): e222685. Epub 2022/10/07.

[191] Salcher-Konrad M, Shi C, Patel D, McDaid D, Astudillo-Garcia CI, Bobrow K, et al. Research evaluating the effectiveness of dementia interventions in low- and middle-income countries: A systematic mapping of 340 randomised controlled trials. Int J Geriatr Psychiatry. 2023;38(7): e5965. Epub 2023/07/11.

[192] ICH. ICH harmonised guideline: general considerations for clinical studies E8(R1). [online]. https://database.ich.org/sites/default/files/E8-R1_Guideline_Step4_2021_1006.pdf (last access 3 Sept 2023): International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use, 2021.

[193] Bothwell LE, Greene JA, Podolsky SH, Jones DS. Assessing the Gold Standard--Lessons from the History of RCTs. N Engl J Med. 2016;374(22): 2175-81. Epub 2016/06/02.

[194] Mehra MR, Desai SS, Ruschitzka F, Patel AN. RETRACTED: Hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: a multinational registry analysis. Lancet. 2020. Epub 2020/05/26.

[195] Prasad V, Cifu A. Medical Reversal: Why We Must Raise the Bar Before Adopting New Technologies. Yale J Biol Med. 2011;84(4): 471-8.

[196] Herrera-Perez D, Haslam A, Crain T, Gill J, Livingston C, Kaestner V, et al. A comprehensive review of randomized clinical trials in three medical journals reveals 396 medical reversals. Elife. 2019;8: e45183. Epub 2019/06/12.

[197] European Commission. Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL laying down Union procedures for the

authorisation and supervision of medicinal products for human use and establishing rules governing the European Medicines Agency, amending Regulation (EC) No 1394/2007 and Regulation (EU) No 536/2014 and repealing Regulation (EC) No 726/2004, Regulation (EC) No 141/2000 and Regulation (EC) No 1901/2006. Brussels: European Commission, 2023.

[198] Mc Cord KA, Ewald H, Agarwal A, Glinz D, Aghlmandi S, Ioannidis JPA, et al. Treatment effects in randomised trials using routinely collected data for outcome assessment versus traditional trials: meta-research study. BMJ. 2021;372: n450. Epub 2021/03/05.

[199] Anglemyer A, Horvath HT, Bero L. Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials. Cochrane Database Syst Rev. 2014;2014(4): MR000034. Epub 2014/05/02.

[200] Ewald H, Ioannidis JPA, Ladanie A, Mc Cord K, Bucher HC, Hemkens LG. Nonrandomized studies using causal-modeling may give different answers than RCTs: a meta-epidemiological study. J Clin Epidemiol. 2020;118: 29-41. Epub 2019/11/11.

[201] Mathes T, Rombey T, Kuss O, Pieper D. No inexplicable disagreements between real-world data-based nonrandomized controlled studies and randomized controlled trials were found. J Clin Epidemiol. 2021;133: 1-13. Epub 2020/12/29.

[202] Moneer O, Daly G, Skydel JJ, Nyhan K, Lurie P, Ross JS, et al. Agreement of treatment effects from observational studies and randomized controlled trials evaluating hydroxychloroquine, lopinavir-ritonavir, or dexamethasone for covid-19: meta-epidemiological study. BMJ. 2022;377: e069400. Epub 2022/05/11.

[203] Savovic J, Jones HE, Altman DG, Harris RJ, Juni P, Pildal J, et al. Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. Ann Intern Med. 2012;157(6): 429-38. Epub 2012/09/05.

[204] Moustgaard H, Jones HE, Savovic J, Clayton GL, Sterne JA, Higgins JP, et al. Ten questions to consider when interpreting results of a meta-epidemiological study-the MetaBLIND study as a case. Res Synth Methods. 2020;11(2): 260-74. Epub 2019/12/19.

[205] Savovic J, Harris RJ, Wood L, Beynon R, Altman D, Als-Nielsen B, et al. Development of a combined database for meta-epidemiological research. Res Synth Methods. 2010;1(3-4): 212-25. Epub 2010/07/01.

[206] IMI GetReal. The RWE Navigator: Generating Real-World Evidence.: NICE, IMI GetReal; 2020 Available from: https://rwe-navigator.eu/use-real-world-evidence/generate-real-world-evidence/ (last access 5 Oct 2023)

[207] Reeves BC, Wells GA, Waddington H. Quasi-experimental study designs series-paper 5: a checklist for classifying studies evaluating the effects on health interventions-a taxonomy without labels. J Clin Epidemiol. 2017;89: 30-42. Epub 2017/03/30.

[208] Grant RL. Converting an odds ratio to a range of plausible relative risks for better communication of research findings. BMJ. 2014;348: f7450. Epub 2014/01/28.

[209] Guyatt GH, Oxman AD, Santesso N, Helfand M, Vist G, Kunz R, et al. GRADE guidelines: 12. Preparing summary of findings tables-binary outcomes. J Clin Epidemiol. 2013;66(2): 158-72. Epub 2012/05/23.

[210] Higgins J, Deeks J. Chapter 7: Selecting studies and collecting data. In: Higgins J, Green S, Eds. Cochrane Handbook for Systematic Reviews of Interventions Version 510 (updated March 2011). Online. Available from www.handbook.cochrane.org: The Cochrane Collaboration; 2011.

[211] Chinn S. A simple method for converting an odds ratio to effect size for use in meta-analysis. Stat Med. 2000;19(22): 3127-31. Epub 2000/12/13.

[212] Page MJ, Higgins JP, Clayton G, Sterne JA, Hrobjartsson A, Savovic J. Empirical Evidence of Study Design Biases in Randomized Trials: Systematic Review of Meta-Epidemiological Studies. PLoS One. 2016;11(7): e0159267. Epub 2016/07/12.

[213] Franklin JM, Dejene S, Huybrechts KF, Wang SV, Kulldorff M, Rothman KJ. A Bias in the Evaluation of Bias Comparing Randomized Trials with Nonexperimental Studies. Epidemiol Methods. 2017;6(1). Epub 2017/12/06.

[214] Borenstein M, Hedges LV, Higgins JP, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. Res Synth Methods. 2010;1(2): 97-111. Epub 2010/04/01.

[215] IntHout J, Ioannidis JP, Borm GF. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. BMC Med Res Methodol. 2014;14: 25. Epub 2014/02/20.

[216] Welton NJ, Ades AE, Carlin JB, Altman DG, Sterne JAC. Models for Potentially Biased Evidence in Meta-Analysis Using Empirically Based Priors. Journal of the Royal Statistical Society Series A (Statistics in Society). 2009;172(1): 119-36.

[217] Savovic J, Turner RM, Mawdsley D, Jones HE, Beynon R, Higgins JPT, et al. Association Between Risk-of-Bias Assessments and Results of Randomized Trials in Cochrane Reviews: The ROBES Meta-Epidemiologic Study. Am J Epidemiol. 2018;187(5): 1113-22. Epub 2017/11/11.

[218] Dahabreh IJ, Sheldrick RC, Paulus JK, Chung M, Varvarigou V, Jafri H, et al. Do observational studies using propensity score methods agree with randomized trials? A systematic comparison of studies on acute coronary syndromes. Eur Heart J. 2012;33(15): 1893-901. Epub 2012/06/20.

[219] Soni PD, Hartman HE, Dess RT, Abugharib A, Allen SG, Feng FY, et al. Comparison of Population-Based Observational Studies With Randomized Trials in Oncology. J Clin Oncol. 2019;37(14): 1209-16. Epub 2019/03/22.

[220] Moustgaard H, Clayton GL, Jones HE, Boutron I, Jorgensen L, Laursen DRT, et al. Impact of blinding on estimated treatment effects in randomised clinical trials: meta-epidemiological study. BMJ. 2020;368: l6802. Epub 2020/01/23.

[221] Lonjon G, Boutron I, Trinquart L, Ahmad N, Aim F, Nizard R, et al. Comparison of treatment effect estimates from prospective nonrandomized studies with propensity score analysis and randomized controlled trials of surgical procedures. Ann Surg. 2014;259(1): 18-25. Epub 2013/10/08.

[222] Kuss O, Legler T, Borgermann J. Treatments effects from randomized trials and propensity score analyses were similar in similar populations in an example from cardiac surgery. J Clin Epidemiol. 2011;64(10): 1076-84. Epub 2011/04/13.

[223] Kaplan RM, Koong AJ, Irvin V. Review of Evidence Supporting 2022 US Food and Drug Administration Drug Approvals. JAMA Netw Open. 2023;6(8): e2327650. Epub 2023/08/08.

[224] Haslam A, Livingston C, Prasad V. Medical Reversals in Family Practice: A Review. Curr Ther Res Clin Exp. 2020;92: 100579. Epub 2020/03/18.

[225] Prasad V, Gall V, Cifu A. The frequency of medical reversal. Arch Intern Med. 2011;171(18): 1675-6. Epub 2011/07/13.

[226] Djurisic S, Rath A, Gaber S, Garattini S, Bertele V, Ngwabyt SN, et al. Barriers to the conduct of randomised clinical trials within all disease areas. Trials. 2017;18(1): 360. Epub 2017/08/03.

[227] Wieseler B, Neyt M, Kaiser T, Hulstaert F, Windeler J. Replacing RCTs with real world data for regulatory decision making: a self-fulfilling prophecy? BMJ. 2023;380: e073100. Epub 2023/03/03.

[228] Higgins JP, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011;343: d5928. Epub 2011/10/20.

[229] Farrah K, Young K, Tunis MC, Zhao L. Risk of bias tools in systematic reviews of health interventions: an analysis of PROSPERO-registered protocols. Syst Rev. 2019;8(1): 280. Epub 2019/11/16.

[230] Salcher-Konrad M, Naci H, Davis C. Approval of Cancer Drugs With Uncertain Therapeutic Value: A Comparison of Regulatory Decisions in Europe and the United States. Milbank Q. 2020;98(4): 1219-56. Epub 2020/10/07.

[231] European Medicines Agency. PRIME: a two-year overview. London [online]. https://www.ema.europa.eu/en/documents/report/prime-two-year-overview_en.pdf (last access 23 Sept 2025): EMA, 2018.

[232] Boucaud-Maitre D, Altman JJ. Do the EMA accelerated assessment procedure and the FDA priority review ensure a therapeutic added value? 2006-2015: a cohort study. Eur J Clin Pharmacol. 2016;72(10): 1275-81. Epub 2016/07/31.

[233] Downing NS, Shah ND, Aminawung JA, Pease AM, Zeitoun JD, Krumholz HM, et al. Postmarket Safety Events Among Novel Therapeutics Approved by the US Food and Drug Administration Between 2001 and 2010. JAMA. 2017;317(18): 1854-63. Epub 2017/05/12.

[234] Zeitoun JD, Lefevre JH, Downing NS, Bergeron H, Ross JS. Regulatory review time and post-market safety events for novel medicines approved by the EMA between 2001 and 2010: a cross-sectional study. Br J Clin Pharmacol. 2015;80(4): 716-26. Epub 2015/03/27.

[235] Grossmann N, Robausch M, Rosian K, Wild C, Simon J. Monitoring evidence on overall survival benefits of anticancer drugs approved by the European Medicines Agency between 2009 and 2015. Eur J Cancer. 2019;110: 1-7. Epub 2019/02/09.

[236] Beakes-Read G, Neisser M, Frey P, Guarducci M. Analysis of FDA's Accelerated Approval Program Performance December 1992-December 2021. Ther Innov Regul Sci. 2022;56(5): 698-703. Epub 2022/07/29.

[237] US Food and Drug Administration. Guidance for Industry: Expedited Programs for Serious Conditions-Drugs and Biologics. Silver Spring, MD [online]. http://www.fda.gov/BiologicsBloodVaccines/GuidanceComplianceRegulatoryInformation/Guidances/default.htm (last access 15 Nov 2022): FDA, 2014.

[238] Hoekman J, Boon WP, Bouvy JC, Ebbers HC, de Jong JP, De Bruin ML. Use of the conditional marketing authorization pathway for oncology medicines in Europe. Clin Pharmacol Ther. 2015;98(5): 534-41. Epub 2015/06/18.

[239] Bloem LT, Schelhaas J, Lopez-Anglada L, Herberts C, van Hennik PB, Tenhunen O. European Conditional Marketing Authorization in a Rapidly Evolving Treatment Landscape: A Comprehensive Study of Anticancer Medicinal Products in 2006-2020. Clin Pharmacol Ther. 2023;114(1): 148-60. Epub 2023/05/02.

[240] European Medicines Agency. Guideline on the scientific application and the practical arrangements necessary to implement Commission Regulation (EC) No 507/2006 on the conditional marketing authorisation for medicinal products for human use falling within the scope of Regulation (EC) No 726/2004. London [online]. https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-scientific-application-and-practical-arrangements-necessary-implement-commission-regulation-ec-no-5072006-conditional-marketing-authorisation-medicinal-products-human-use-falling_en.pdf (last access 23 Sept 2025): EMA, 2016.

[241] European Medicines Agency. European Medicines Agency post-authorisation procedural advice for users of the centralised procedure. EMEA-H-19984/03 Rev. 82. Amsterdam [online]. https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/european-medicines-agency-post-authorisation-procedural-advice-users-centralised-procedure_en.pdf (last access 23 Sept 2025): EMA, 2019.

[242] US Food and Drug Administration. Guidance for Industry: Postmarketing Studies and Clinical Trials — Implementation of Section 505(o)(3) of the Federal Food, Drug, and Cosmetic Act. Silver Spring, MD [online]. https://www.fda.gov/media/116878/download (last access 15 Nov 2022): FDA, 2011.

[243] Vokinger KN, Kesselheim AS. Application of orphan drug designation to cancer treatments (2008-2017): a comprehensive and comparative analysis of the USA and EU. BMJ Open. 2019;9(10): e028634. Epub 2019/10/12.

[244] Martinalbo J, Bowen D, Camarero J, Chapelin M, Demolis P, Foggi P, et al. Early market access of cancer drugs in the EU. Ann Oncol. 2016;27(1): 96-105. Epub 2015/10/22.

[245] Johnson JR, Ning YM, Farrell A, Justice R, Keegan P, Pazdur R. Accelerated approval of oncology products: The food and drug administration experience. Journal of the National Cancer Institute. 2011;103: 636-44.

[246] Grossmann N, Wild C. Between January 2009 and April 2016, 134 novel anticancer therapies were approved: what is the level of knowledge concerning the clinical benefit at the time of approval? ESMO Open. 2016;1(6): e000125. Epub 2017/08/30.

[247] Roberts SA, Allen JD, Sigal EV. Despite criticism of the FDA review process, new cancer drugs reach patients sooner in the United States than in Europe. Health Aff (Millwood). 2011;30(7): 1375-81. Epub 2011/06/18.

[248] Beishon M. Approval rating: how do the EMA and FDA compare? Cancer World. 2014;1: 12-7.

[249] Trotta F, Leufkens HG, Schellens JH, Laing R, Tafuri G. Evaluation of oncology drugs at the European Medicines Agency and US Food and Drug Administration: when differences have an impact on clinical practice. J Clin Oncol. 2011;29(16): 2266-72. Epub 2011/05/04.

[250] Tafuri G, Stolk P, Trotta F, Putzeist M, Leufkens HG, Laing RO, et al. How do the EMA and FDA decide which anticancer drugs make it to the market? A comparative qualitative study on decision makers' views. Ann Oncol. 2014;25(1): 265-9. Epub 2013/12/21.

[251] Osterborg A, Udvardy M, Zaritskey A, Andersson PO, Grosicki S, Mazur G, et al. Phase III, randomized study of ofatumumab versus physicians' choice of therapy and standard versus extended-length ofatumumab in patients with bulky fludarabine-refractory chronic lymphocytic leukemia. Leuk Lymphoma. 2016;57(9): 2037-46. Epub 2016/01/20.

[252] Hillmen P, Robak T, Janssens A, Babu KG, Kloczko J, Grosicki S, et al. Chlorambucil plus ofatumumab versus chlorambucil alone in previously untreated patients with chronic lymphocytic leukaemia (COMPLEMENT 1): a randomised, multicentre, open-label phase 3 trial. Lancet. 2015;385(9980): 1873-83. Epub 2015/04/18.

[253] Gelmon KA, Boyle FM, Kaufman B, Huntsman DG, Manikhas A, Di Leo A, et al. Lapatinib or Trastuzumab Plus Taxane Therapy for Human Epidermal Growth Factor Receptor 2-Positive Advanced Breast Cancer: Final Results of NCIC CTG MA.31. J Clin Oncol. 2015;33(14): 1574-83. Epub 2015/03/18.

[254] European Medicines Agency. Committee for medicinal products for human use (CHMP): Minutes of the meeting on 15-18 October 2018. London [online]. https://www.ema.europa.eu/documents/minutes/minutes-chmp-meeting-15-18-october-2018_en.pdf (last access 23 Sept 2025): EMA, 2018.

[255] Connors JM, Jurczak W, Straus DJ, Ansell SM, Kim WS, Gallamini A, et al. Brentuximab Vedotin with Chemotherapy for Stage III or IV Hodgkin's Lymphoma. N Engl J Med. 2018;378(4): 331-44. Epub 2017/12/12.

[256] European Medicines Agency. Iclusig EPAR assessment report EMEA/H/C/002695. London, 2013.

[257] F. D. A. Center for Drug Evaluation Research. Cross discipline team leader review: Iclusig/ponatinib NDA 203469. 2012.

[258] Lipton JH, Chuah C, Guerci-Bresler A, Rosti G, Simpson D, Assouline S, et al. Ponatinib versus imatinib for newly diagnosed chronic myeloid leukaemia: an international, randomised, open-label, phase 3 trial. Lancet Oncol. 2016;17(5): 612-21. Epub 2016/04/17.

[259] Prasad V, Mailankody S. The accelerated approval of oncologic drugs: lessons from ponatinib. JAMA. 2014;311(4): 353-4. Epub 2014/01/23.

[260] EMA Pharmacovigilance Risk Assessment Committee (PRAC). Iclusig Assessment Report: Review under Article 20 of Regulation (EC) No 726/2004. London, 2014.

[261] Cortes JE, Kim D-W, Pinilla-Ibarz J, Le Coutre P, Paquette R, Chuah C, et al. Long-Term Follow-up of Ponatinib Efficacy and Safety in the Phase 2 PACE Trial. Blood. 2014;124(21).

[262] European Medicines Agency. Bosulif EPAR Assessment Report EMEA/H/C/002237. London: EMA, 2013.

[263] F. D. A. Center for Drug Evaluation Research. Memorandum of Meeting Minutes: Type B Meeting for IND 68,268 on 23 Oct 2007 between FDA Division of Oncology Products and Wyeth. [online]. https://www.accessdata.fda.gov/drugsatfda_docs/nda/2012/203341Orig1s000Admincorres.pdf (last access 17 June 2025), 2007.

[264] F. D. A. Center for Drug Evaluation Research. Cross Discipline Team Leader Review: Bosulif/Bosutinib NDA 203341. [online]. https://www.accessdata.fda.gov/drugsatfda_docs/nda/2012/203341Orig1s000CrossR.pdf (last access 17 June 2025), 2012.

[265] F. D. A. Center for Drug Evaluation Research. Summary Review: Bosulif/Bosutinib NDA 203341. [online]. https://www.accessdata.fda.gov/drugsatfda_docs/nda/2012/203341Orig1s000SumR.pdf (last access 25 June 2025), 2012.

[266] F. D. A. Center for Drug Evaluation Research. Statistical Review: Bosulif/Bosutinib NDA 203341. [online]. https://www.accessdata.fda.gov/drugsatfda_docs/nda/2012/203341Orig1s000StatR.pdf (last access 25 June 2025), 2011.

[267] F. D. A. Center for Drug Evaluation Research. Summary review: Folotyn/pralatrexate injection NDA 022468. 2009.

[268] European Medicines Agency. Folotyn EPAR assessment report EMEA/H/C/002096. London: EMA, 2012.

[269] Skydel JJ, Luxkaranayagam AT, Dhruva SS, Ross JS, Wallach JD. Analysis of Postapproval Clinical Trials of Therapeutics Approved by the US Food and Drug Administration Without Clinical Postmarketing Requirements or Commitments. JAMA Netw Open. 2019;2(5): e193410. Epub 2019/05/11.

[270] Naci H, Salcher-Konrad M, Kesselheim AS, Wieseler B, Rochaix L, Redberg RF, et al. Generating comparative evidence on new drugs and devices before approval. Lancet. 2020;395(10228): 986-97. Epub 2020/03/23.

[271] Bloem LT, Mantel-Teeuwisse AK, Leufkens HGM, De Bruin ML, Klungel OH, Hoekman J. Postauthorization Changes to Specific Obligations of Conditionally Authorized Medicines in the European Union: A Retrospective Cohort Study. Clin Pharmacol Ther. 2019;105(2): 426-35. Epub 2018/07/04.

[272] Schuenemann H, Brozek J, Guyatt G, Oxman A. Handbook for Grading the Quality of Evidence and the Strength of Recommendations Using the GRADE Approach. [online]. https://gdt.gradepro.org/app/handbook/handbook.html (last access 7 Nov 2018), 2013.

[273] European Medicines Agency. ICH Topic E 9 Statistical Principles for Clinical Trials: Step 5 Note for guidance on statistical principles for clinical trials

(CPMP/ICH/363/96). London [online]. https://www.ema.europa.eu/documents/scientific-guideline/ich-e-9-statistical-principles-clinical-trials-step-5_en.pdf (last access 15 Sept 2023): EMA, 1998.

[274] European Medicines Agency. Scientific guidance on post-authorisation efficacy studies. London [online]. https://www.ema.europa.eu/en/documents/scientific-guideline/scientific-guidance-post-authorisation-efficacy-studies-first-version_en.pdf (last access 23 Sept 2025): EMA, 2016.

[275] Kovic B, Jin X, Kennedy SA, Hylands M, Pedziwiatr M, Kuriyama A, et al. Evaluating Progression-Free Survival as a Surrogate Outcome for Health-Related Quality of Life in Oncology: A Systematic Review and Quantitative Analysis. JAMA Intern Med. 2018;178(12): 1586-96. Epub 2018/10/05.

[276] Wieseler B, McGauran N, Kaiser T. New drugs: where did we go wrong and what can we do better? BMJ. 2019;366: l4340. Epub 2019/07/12.

[277] Silverman E. Public Citizen urges FDA to withdraw drug for preventing premature birth. Pharmalot; 2019.

[278] Annex I to the DIRECTIVE 2001/83/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 6 November 2001 on the Community code relating to medicinal products for human use.

[279] Salcher-Konrad M, Nguyen M, Savovic J, Higgins JPT, Naci H. Treatment Effects in Randomized and Nonrandomized Studies of Pharmacological Interventions: A Meta-Analysis. JAMA Netw Open. 2024;7(9): e2436230. Epub 2024/09/27.

[280] Agrawal S, Arora S, Vallejo JJ, Gwise T, Chuk MK, Amiri-Kordestani L, et al. Use of single-arm trials to support malignant hematology and oncology drug and biologic approvals: A 20-year FDA experience. Journal of Clinical Oncology. 2021;39(15_suppl): e13572-e.

[281] Simon R, Blumenthal GM, Rothenberg ML, Sommer J, Roberts SA, Armstrong DK, et al. The role of nonrandomized trials in the evaluation of oncology drugs. Clin Pharmacol Ther. 2015;97(5): 502-7. Epub 2015/02/14.

[282] Blumenthal GM, Pazdur R. Response Rate as an Approval End Point in Oncology: Back to the Future. JAMA Oncol. 2016;2(6): 780-1. Epub 2016/02/26.

[283] Glasziou P, Chalmers I, Rawlins M, McCulloch P. When are randomised trials unnecessary? Picking signal from noise. BMJ. 2007;334(7589): 349-51. Epub 2007/02/17.

[284] Martinalbo J, Camarero J, Delgado-Charro B, Démolis P, Ersbøll J, Foggi P, et al. Single-arm trials for cancer drug approval and patient access. Annals of Oncology. 2016;27: vi584.

[285] European Society for Medical Oncology. ESMO 2016 Press Release: Single-Arm Trials Improve Early Access to Rare Cancer Drugs. Lugano, Copenhagen [online]. https://www.esmo.org/meeting-calendar/past-meetings/esmo-2016-congress/press-media/single-arm-trials-improve-early-access-to-rare-cancer-drugs (last access 16 Nov 2022): ESMO, 2016.

[286] European Medicines Agency. Guideline on clinical trials in small populations. London [online]. https://www.ema.europa.eu/en/documents/scientific-

guideline/guideline-clinical-trials-small-populations_en.pdf (last access 24 June 2025): EMA, 2006.

[287] Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024;74(3): 229-63. Epub 2024/04/04.

[288] Kranz P, Egger G, Kaiser T, Lotz F, Ünal C, Waffenschmidt S. Working Paper GA21-01: Evidence on orphan drugs. Cologne [online]. https://www.iqwig.de/download/ga21-01_evidence-on-orphan-drugs_working-paper_v1-0.pdf (last access 9 Jan 2025): IQWiG - Institute for Quality and Efficiency in Health Care, 2021.

[289] Sakpal TV. Sample size estimation in clinical trial. Perspect Clin Res. 2010;1(2): 67-9. Epub 2010/04/01.

[290] Kirby S, Burke J, Chuang-Stein C, Sin C. Discounting phase 2 results when planning phase 3 clinical trials. Pharm Stat. 2012;11(5): 373-85. Epub 2012/05/30.

[291] Wang SJ, Hung HM, O'Neill RT. Adapting the sample size planning of a phase III trial based on phase II data. Pharm Stat. 2006;5(2): 85-97. Epub 2006/11/04.

[292] Sim J. Should treatment effects be estimated in pilot and feasibility studies? Pilot Feasibility Stud. 2019;5: 107. Epub 2019/09/06.

[293] Winstone J, Chadda S, Ralston S, Sajosi P. Review and comparison of clinical evidence submitted to support European Medicines Agency market authorization of orphan-designated oncological treatments. Orphanet J Rare Dis. 2015;10: 139. Epub 2015/10/30.

[294] van Hattem CC, de Jong AJ, de Groot JS, Hoekman J, Broekman KE, Sonke GS, et al. Factors affecting the feasibility of post-authorisation RCTs for conditionally authorised anticancer medicines: a multistakeholder perspective from a qualitative focus group study. BMJ Open. 2024;14(11): e084483. Epub 2024/11/13.

[295] Wallach JD, Ramachandran R, Bruckner T, Ross JS. Comparison of Duration of Postapproval vs Pivotal Trials for Therapeutic Agents Granted US Food and Drug Administration Accelerated Approval, 2009-2018. JAMA Netw Open. 2021;4(11): e2133601. Epub 2021/11/10.

[296] Fernandez Lynch H, Sachs RE, Lee S, Herder M, Ross JS, Ramachandran R. Extending the US Food and Drug Administration's Postmarket Authorities. JAMA Health Forum. 2023;4(6): e231313. Epub 2023/06/09.

[297] Beaver JA, Pazdur R. "Dangling" Accelerated Approvals in Oncology. N Engl J Med. 2021;384(18): e68. Epub 2021/04/22.

[298] Kim C, Prasad V. Cancer Drugs Approved on the Basis of a Surrogate End Point and Subsequent Overall Survival: An Analysis of 5 Years of US Food and Drug Administration Approvals. JAMA Intern Med. 2015;175(12): 1992-4. Epub 2015/10/27.

[299] Tsimberidou AM, Braiteh F, Stewart DJ, Kurzrock R. Ultimate fate of oncology drugs approved by the us food and drug administration without a randomized Trial. J Clin Oncol. 2009;27(36): 6243-50. Epub 2009/10/15.

[300] Wong AK, Mooghali M, Ramachandran R, Ross JS, Wallach JD. Use of Expedited Regulatory Programs and Clinical Development Times for FDA-Approved Novel Therapeutics. JAMA Netw Open. 2023;6(8): e2331753. Epub 2023/08/31.

[301] Demirci E, Omes-Smit G, Zwiers A. Clinical development time is shorter for new anticancer drugs approved via accelerated approval in the US or via conditional approval in the EU. Clin Transl Sci. 2023;16(7): 1127-33. Epub 2023/04/05.

[302] Brinkhuis F, Goettsch WG, Mantel-Teeuwisse AK, Bloem LT. Added benefit and revenues of oncology drugs approved by the European Medicines Agency between 1995 and 2020: retrospective cohort study. BMJ. 2024;384: e077391. Epub 2024/02/29.

[303] Moore TJ, Zhang H, Anderson G, Alexander GC. Estimated Costs of Pivotal Trials for Novel Therapeutic Agents Approved by the US Food and Drug Administration, 2015-2016. JAMA Intern Med. 2018;178(11): 1451-7. Epub 2018/09/29.

[304] Tibau A, Molto C, Borrell M, Del Paggio JC, Barnadas A, Booth CM, et al. Magnitude of Clinical Benefit of Cancer Drugs Approved by the US Food and Drug Administration Based on Single-Arm Trials. JAMA Oncology. 2018;4: 1610-1.

[305] Angelis A, Lange A, Kanavos P. Using health technology assessment to assess the value of new medicines: results of a systematic review and expert consultation across eight European countries. Eur J Health Econ. 2018;19(1): 123-52. Epub 2017/03/18.

[306] Claessens Z, Fieuws S, Daems J, Barbier L, Huys I. Do European regulatory measures accelerate national market access in Belgium? A retrospective analysis of medicines centrally authorised between 2015 and 2020. BMJ Open. 2025;15(1): e091361. Epub 2025/01/10.

[307] Mills M. HTA Barriers for Conditional Approval Drugs. PharmacoEconomics. 2023;41(5): 529-45. Epub 2023/02/24.

[308] Gallo V, Alessi E, Montilla S, Altamura G, Traversa G, Trotta F. The timelines for the price and reimbursement authorization in Italy 2018-2020. Front Med (Lausanne). 2022;9: 1055359. Epub 2023/01/10.

[309] Vreman RA, Bouvy JC, Bloem LT, Hovels AM, Mantel-Teeuwisse AK, Leufkens HGM, et al. Weighing of Evidence by Health Technology Assessment Bodies: Retrospective Study of Reimbursement Recommendations for Conditionally Approved Drugs. Clin Pharmacol Ther. 2019;105(3): 684-91. Epub 2018/10/10.

[310] Jansen JP, Naci H. Is network meta-analysis as valid as standard pairwise meta-analysis? It all depends on the distribution of effect modifiers. BMC Med. 2013;11: 159. Epub 2013/07/06.

[311] Wallach JD, Yoon S, Doernberg H, Glick LR, Ciani O, Taylor RS, et al. Associations Between Surrogate Markers and Clinical Outcomes for Nononcologic Chronic Disease Treatments. JAMA. 2024;331(19): 1646-54. Epub 2024/04/22.

[312] Koong AJ, Irvin VL, Narayan A, Song S, Kaplan RM. Evidence available and used by the Food and Drug Administration for the approval of orphan and nonorphan drugs. Health Aff Sch. 2025;3(4): qxaf057. Epub 2025/04/07.

[313] Eichler HG, Bloechl-Daum B, Abadie E, Barnett D, Konig F, Pearson S. Relative efficacy of drugs: an emerging issue between regulatory agencies and third-party payers. Nat Rev Drug Discov. 2010;9(4): 277-91. Epub 2010/02/27.

[314] European Medicines Agency. Parallel scientific advice and special development aspects or product types. 2025 Available from: https://www.ema.europa.eu/en/human-regulatory-overview/research-development/scientific-advice-protocol-assistance/parallel-scientific-advice-special-development-aspects-or-product-types (last access 17 Sept 2025)

[315] Hofer MP, Jakobsson C, Zafiropoulos N, Vamvakas S, Vetter T, Regnstrom J, et al. Regulatory watch: Impact of scientific advice from the European Medicines Agency. Nat Rev Drug Discov. 2015;14(5): 302-3. Epub 2015/04/18.

[316] Jazowski SA, Vaidya AU, Donohue JM, Dusetzina SB, Sachs RE. Time to Confirmatory Study Initiation After Accelerated Approval of Cancer and Noncancer Drugs in the US. JAMA Intern Med. 2023;183(7): 737-9. Epub 2023/05/22.

[317] Vokinger KN, Li G, Wouters OJ. The Rise of Drug Innovation in China - Implications for Patient Access in the United States and Globally. N Engl J Med. 2025;393(9): 839-41. Epub 2025/09/01.

[318] Tafuri G, Pagnini M, Moseley J, Massari M, Petavy F, Behring A, et al. How aligned are the perspectives of EU regulators and HTA bodies? A comparative analysis of regulatory-HTA parallel scientific advice. Br J Clin Pharmacol. 2016;82(4): 965-73. Epub 2016/06/02.

[319] Prilla S, Groeneveld S, Pacurariu A, Restrepo-Mendez MC, Verpillat P, Torre C, et al. Real-World Evidence to Support EU Regulatory Decision Making-Results From a Pilot of Regulatory Use Cases. Clin Pharmacol Ther. 2024;116(5): 1188-97. Epub 2024/07/04.

[320] Sculpher M, Claxton K, Pearson SD. Developing a Value Framework: The Need to Reflect the Opportunity Costs of Funding Decisions. Value Health. 2017;20(2): 234-9. Epub 2017/02/27.

[321] Kater AP, Harrup R, Kipps TJ, Eichhorst B, Owen CJ, Assouline S, et al. The MURANO study: final analysis and retreatment/crossover substudy results of VenR for patients with relapsed/refractory CLL. Blood. 2025;145(23): 2733-45. Epub 2025/02/26.

[322] Naci H, Murphy P, Woods B, Lomas J, Wei J, Papanicolas I. Population-health impact of new drugs recommended by the National Institute for Health and Care Excellence in England during 2000-20: a retrospective analysis. Lancet. 2025;405(10472): 50-60. Epub 2024/12/16.

Appendix

Appendix 1: Review of regulatory guidance on non-randomised studies

Regulatory guidance on single-arm trials

Methodological challenges related to causal inference from single-arm trials are reflected in guidance issued by regulatory bodies in the EU and US. The EMA highlights that, even in situations where RCTs are difficult to conduct (e.g. due to rarity of disease or the lack of alternative treatments), they are preferred to other study designs to their methodological advantages.[16] In a reflection paper on the use of single-arm trials for regulatory approval, the EMA stresses that causal interpretation of an effect estimate obtained from a single-arm trial is only possible when the outcome could not be observed for any enrolled participant without the treatment being effective.[23] In addition to the selection of an appropriate endpoint, several other measures should be implemented to reduce bias in single-arm trials, including blinded assessment of endpoints, avoiding missing data, adhering to the protocol and pre-planning the statistical analysis, defining the target population independently of disease severity, and defining inclusion/exclusion criteria matching a hypothetical control group, among others.[23]

The FDA, in its guidance on accelerated approvals in oncology and for biologic medicines, also refers to methodological challenges of single-arm trials, including the typically small size of safety databases, limits to interpretation of time-to-event endpoints such as survival, small effect sizes not generally being likely to predict clinical benefit, identification of the contributions of individual agents in combination therapies, and lack of comparability of results across trials for contextualisation.[19] Accordingly, RCTs are preferred as pivotal trials, but regulators acknowledge that single-arm trials may be appropriate for approval in a specific clinical and regulatory context. Specifically, the FDA mentions concerns about the feasibility of conducting a RCT,[19] e.g. due to a small number of eligible patients.[18] Both FDA and EMA refer to the need for a clearly established natural course of disease, where the outcome of the disease without treatment is predictable, as a pre-requisite for making causal inference about treatment effects from studies without a concurrent control.[16, 18, 23]

Regulatory guidance on external controls

Regulators are well aware of the limitations of external controls. Due to the severe risk of bias, externally controlled studies should only be used in specific situations according to regulatory guidance documents. In harmonised guidance, the EMA and FDA state the following:

> *"[I]nability to control bias restricts use of the external control design to situations in which the effect of treatment is dramatic and the usual course of the disease highly predictable. In addition, use of external controls should be limited to cases in which the endpoints are objective and the impact of baseline and treatment variables on the endpoint is well characterized."* [16, p26, 162, p27]

For the analysis of data from non-randomised studies, regulatory guidance documents do not specify appropriate analytical methods and instead focus on general principles that should be adhered to when using external controls.[16, 162] To address potential bias and increase the validity of effect estimates obtained from externally controlled studies, regulators recommend that detailed information – ideally, participant-level data – should be available for external controls. External controls and their treatment setting should further resemble the interventional study as closely as possible, except for the new treatment being studied. EMA and FDA further stress the need for pre-specification of the external control group prior to conducting the pivotal interventional study, including specifying any matching or adjustments prior to conducting an externally controlled study, and emphasise that sensitivity analyses may be needed in particular for observational studies.[16, 17, 162]

Beyond these general principles, regulators may provide specific guidance on external control groups and analytical questions for individual marketing authorisation applications through scientific advice provided to medicines developers prior to submission for regulatory approval. The EMA reported that there were 26 requests for scientific advice in 2024 that related to real-world data, including some relating to the use of historical controls as reference for defining endpoint thresholds and for an externally controlled study.[40]

Both EMA and FDA have published documents outlining their thinking about planning, conducting, analysing, and reporting observational studies using real-world

evidence for regulatory purposes (a reflection paper for the EU and draft guidance for the US).[129, 130] These documents do not (yet) constitute formal guidance but they clearly show the interest of the regulatory agencies in making causal inferences from observational data about the clinical benefits of medicines. The FDA explicitly states that its draft guidance "addresses the growing interest in the potential use of non-interventional studies to support the demonstration of the effectiveness of a drug."[130, p3]

A key requirement for adding credibility to non-interventional studies is transparency, as reflected in regulatory guidance documents as well as international standards for conducting, analysing, and reporting observational studies. This can include a pre-specified protocol including planned statistical analyses and transparent reporting of results, and may even include registration of observational studies (similar to clinical trial registration).[39, 129, 130, 168, 169] Regulators also recommend early engagement to discuss the relevance and feasibility of observational studies for regulatory purposes.[52, 130]

In its reflection paper on the use of real-world data in non-interventional studies, the EMA recommends applying the target trial emulation approach when designing non-interventional studies with a causal question.[129] This approach includes first specifying a hypothetical trial that would address the causal question (the target trial) and to then design a non-randomised study which emulates the target trial as closely as possible for each of the specified design elements.[49] Additional measures to address risk of bias still need to be applied in observational studies, even when using the target trial emulation approach. The EMA reflection paper discusses the different risks of bias in non-interventional studies and methods to mitigate them, including selection bias, information bias, time-related bias, and confounding.[129]

Prior to its reflection paper on non-intervention studies, the EMA had published a guideline on observational studies based on patient registries.[52] This guidance emerged from a multi-stakeholder consultation process driven by a desire to better capitalise on the data available in patient registries.[53] Note that the guidance relates to patient registries only and excludes so-called product registries (i.e. the guidance focuses on registries that are centred around individual patients or conditions, rather than individual medicinal products). The guidance clearly distinguishes between

patient registries (an organised system for data collection) and registry-based studies (which use data from patient registries) and provides recommendations on each step for conducting registry-based studies, including study planning, development of a study protocol, definition of the study population and how to handle informed consent and data protection, data collection, data quality management, data analysis, and reporting.[52]

The FDA's draft guidance on use of real-world data for observational studies addresses similar issues as the EMA guidance but is more prescriptive. The document lists specific elements that medicines developers should address when proposing an observational study to support regulatory approval. This includes items related to study design, data sources, and analysis.[130]

Regulatory guidance on observational cohort studies

Both EMA and FDA have published documents outlining their thinking about planning, conducting, analysing, and reporting observational studies using real-world evidence for regulatory purposes (a reflection paper for the EU and draft guidance for the US).[129, 130] These documents do not (yet) constitute formal guidance but they clearly show the interest of the regulatory agencies in making causal inferences from observational data about the clinical benefits of medicines. The FDA explicitly states that its draft guidance "addresses the growing interest in the potential use of non-interventional studies to support the demonstration of the effectiveness of a drug."[130, p3]

A key requirement for adding credibility to non-interventional studies is transparency, as reflected in regulatory guidance documents as well as international standards for conducting, analysing, and reporting observational studies. This can include a pre-specified protocol including planned statistical analyses and transparent reporting of results, and may even include registration of observational studies (similar to clinical trial registration).[39, 129, 130, 168, 169] Regulators also recommend early engagement to discuss the relevance and feasibility of observational studies for regulatory purposes.[52, 130]

In its reflection paper on the use of real-world data in non-interventional studies, the EMA recommends applying the target trial emulation approach when designing non-interventional studies with a causal question.[129] This approach includes first

specifying a hypothetical trial that would address the causal question (the target trial) and to then design a non-randomised study which emulates the target trial as closely as possible for each of the specified design elements.[49] Additional measures to address risk of bias still need to be applied in observational studies, even when using the target trial emulation approach. The EMA reflection paper discusses the different risks of bias in non-interventional studies and methods to mitigate them, including selection bias, information bias, time-related bias, and confounding.[129]

Prior to its reflection paper on non-intervention studies, the EMA had published a guideline on observational studies based on patient registries.[52] This guidance emerged from a multi-stakeholder consultation process driven by a desire to better capitalise on the data available in patient registries.[53] Note that the guidance relates to patient registries only and excludes so-called product registries (i.e. the guidance focuses on registries that are centred around individual patients or conditions, rather than individual medicinal products). The guidance clearly distinguishes between patient registries (an organised system for data collection) and registry-based studies (which use data from patient registries) and provides recommendations on each step for conducting registry-based studies, including study planning, development of a study protocol, definition of the study population and how to handle informed consent and data protection, data collection, data quality management, data analysis, and reporting.[52]

The FDA's draft guidance on use of real-world data for observational studies addresses similar issues as the EMA guidance but is more prescriptive. The document lists specific elements that medicines developers should address when proposing an observational study to support regulatory approval. This includes items related to study design, data sources, and analysis.[130]

Regulatory guidance on real-world data

Worldwide, regulators are increasingly interested in leveraging real-world data throughout the product life cycle, i.e. going beyond the traditional use for pharmacovigilance and other post-marketing studies, and also including market approval for new products and for extensions of already authorised products. Accordingly, regulators have started developing frameworks for how to use real-world

data. In a review of the current state of guidance development, Burns et al. mapped efforts from regulatory bodies across the world in three phases.[26] First, regulators need to develop frameworks to guide the intended use of real-world data and real-world evidence. These frameworks are then operationalised through guidance on real-world data quality and standards. Finally, regulators can provide guidance on study methods for analysing real-world data. Regulators have started developing frameworks and guidance across these three phases, with EMA and FDA taking leading roles, although there remains scope for more detailed and specific guidance on real-world data.

As discussed in section 2.4.2, real-world data can play an important role as source for external controls and thus feature in initial approval decisions. Use of such data is in line with the description of external controls in harmonised guidance, although the term "real-world data" was not in use when these guidance documents were developed. There are currently no specific guidelines on real-world data as external controls. However, EMA and FDA have published frameworks for the use of real-world data in regulatory decision making,[35, 39, 171] as have other regulatory bodies.[26] The EMA and FDA frameworks lay out plans for developing guidance on real-world data in relation to specific topics, such as data quality, real-world data in non-interventional (observational) studies, and real-world data in clinical trials.

A common feature of regulatory guidance regarding real-world data is that their planned use should be discussed at an early stage between the medicine developer and the regulatory body, e.g. through scientific advice. EMA and FDA also discuss the need for a feasibility assessment of whether the research question at hand can be addressed through real-world data.[129, 172] A key feature of regulatory guidance on real-world data is to ensure high quality of the data. Both EMA and FDA have published frameworks on data quality and their application to real-world data, including assessments of data reliability (i.e. are the data correct?), extensiveness (are data sufficient?), coherence (are data homogeneous?), timeliness (are data timely?), and relevance (are the data right for the question?).[173, 174]

Published FDA guidance on the use of real-world data to support regulatory decision making does not refer to specific situations when such data would be considered acceptable for approval decisions.[172] However, in its framework for a real-

world evidence programme, the FDA lists three specific use cases for which it aims to explore the potential of real-world data to inform regulatory decisions regarding effectiveness.[39] These include making changes to the label (including, among others, adding a new indication), adding a new population, and adding safety or effectiveness information. These use cases to further explore the use of real-world data therefore all relate to supplemental indications or changes to existing approvals. However, the FDA has also used historical controls based on real-world data to support initial approvals of new medicines. In its real-world evidence framework, these are described as limited instances in oncology and rare diseases.[39]

The framework also refers to recent efforts to replicate results of RCTs in observational studies and states the FDA's interest in using observational studies to contribute evidence on a product's effectiveness, theoretically widening the scope of the use of real-world data for regulatory decisions.[39] Senior FDA officials have discussed the use of real-world evidence to establish effectiveness as an "emerging science".[34] The FDA's real-world evidence framework emphasizes the importance of transparency in the planning, conduct, and reporting of observational studies.

The EMA has laid out plans for developing guidance on real-world data quality, its use in non-interventional and clinical studies, and on submissions of real-world data for regulatory decisions.[171] As part of this development programme, a reflection paper on real-world data in non-interventional studies was published in 2025,[129] and work on a concept for using real-world data as source for external controls was included in the 2025-2027 work plan of the methodological working party.[171] The reflection paper does not provide guidance on appropriate use cases of non-interventional studies using real-world data for regulatory approval, but it lists examples of how such data have been used in the past. For initial approval, this was limited to descriptive purposes, including studies to characterise disease epidemiology, clinical practice, and drug utilisation patterns, to assess feasibility of post-marketing studies, and to compare patient characteristics in clinical trials to real-world populations.[129] In the post-marketing phase, uses of real-world data in non-interventional studies included safety and effectiveness assessments. The reflection paper provides recommendations for conducting non-interventional studies to address causal questions, demonstrating a clear interest in use of real-world data beyond traditional monitoring activities.

Appendix 2: Search strategy for chapter 3

The search strategy for meta-analyses including both randomised controlled trials (RCTs) and non-randomised studies (NRS) consisted of three parts:

(1) a database search in MEDLINE for meta-analyses including both RCTs and NRS,
(2) a review of all systematic reviews indexed in the Cochrane Database of Systematic Reviews that included both RCTs and NRS,
(3) a review of existing meta-epidemiological studies comparing RCTs and NRS.

Search terms used and number of records identified for each of these three parts are presented below.

Part 1: Search for systematic reviews including randomised and non-randomised studies (MEDLINE via PubMed)

Part 1 of the search strategy involved searching MEDLINE (via PubMed) for systematic reviews that included both RCTs and non-randomised studies and that included meta-analysis. The search was conducted on 8 October 2018. The database was searched from March 2000 onwards (the date when Ioannidis et al., 2001 last updated their search of MEDLINE using the same search strategy). Compared to the search used by Ioannidis et al. (2001), search terms were added to identify reviews of pharmacological therapies (medicines) only.

Table A 1: Search strategy for systematic reviews including both randomised and non-randomised studies, published since 2000 (MEDLINE via PubMed)

Search #	Search terms	Number of records
#1	"Observational Studies as Topic"[Mesh]	3,058
#2	"Cohort Studies"[Mesh]	1,750,345
#3	"Controlled Before-After Studies"[Mesh]	330
#4	"Cross-Sectional Studies"[Mesh]	267,550
#5	"Historically Controlled Study"[Mesh]	138
#6	"Interrupted Time Series Analysis"[Mesh]	433
#7	"Case-Control Studies"[Mesh]	921,604
#8	"Non-Randomized Controlled Trials as Topic"[Mesh]	357
#9	"Propensity Score"[Mesh]	5,576
#10	#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9	2,152,490
#11	"Randomized Controlled Trials as Topic"[Mesh]	118,781

#12	("Meta-Analysis as Topic"[Mesh] OR "Network Meta-Analysis"[Mesh] OR "Meta-Analysis" [Publication Type])	104,595
#13	(pharmacologic therapy[MeSH Terms]) OR drug*[Text Word] OR medicine*[Text Word])	6,339,992
#14	#10 AND #11 AND #12 AND #13	1,303
#15	Letter[ptyp] OR Editorial[ptyp] OR Comment[ptyp]	
#16	#14 NOT #15	1,266
	Limit to humans only	1,260
	Limit to hits from March 2000-now	1,156

Part 2: Search for systematic reviews including randomised and non-randomised studies (Cochrane Database of Systematic Reviews)

Part 2 involved searching the Cochrane Database of Systematic Reviews to identify systematic reviews that included both RCTs and NRS. The search was conducted on 8 October 2018.

Table A 2: Search strategy for Cochrane systematic reviews including both randomised and non-randomised studies (Cochrane Database of Systematic Reviews)

Search #	Search terms	Number of records
#1	Observational:ti,ab,kw	15,169
#2	MeSH descriptor: [Observational Studies as Topic] explode all trees	58
#3	non-randomized:ti,ab,kw	2,251
#4	MeSH descriptor: [Non-Randomized Controlled Trials as Topic] explode all trees	25
#3	quasi:ti,ab,kw	5,146
#5	cohort:ti,ab,kw	42,795
#6	MeSH descriptor: [Cohort Studies as Topic] explode all trees	142,696
#7	before-after:ti,ab,kw	1,035
#8	MeSH descriptor: [Controlled Before-After Studies] explode all trees	39
#9	Cross-sectional:ti,ab,kw	13,194
#10	MeSH descriptor: [Cross-Sectional Studies] explode all trees	4,628
#11	historica*:ti,ab,kw	3,879

#12	MeSH descriptor: [Historically Controlled Study] explode all trees	21
#13	Interrupted time series:ti,ab,kw	459
#14	MeSH descriptor: [Interrupted Time Series Analysis] explode all trees	23
#15	Case-control:ti,ab,kw	10,171
#16	MeSH descriptor: [Case-Control Studies] explode all trees	15,067
#17	Match*:ti,ab,kw	34,468
#18	Propensity:ti,ab,kw	2,845
#19	MeSH descriptor: [Propensity Score] explode all trees	199
#20	#1 or #2 or #3 or #4 or #5 or #6 or #7 or #8 or #9 or #10 or #11 or #12 or #13 or #14 or #15 or #16 or #17 or #18 or #19	230,725
#21	randomized:ti,ab,kw or rct:ti,ab,kw or "randomized controlled clinical trial":ti,ab,kw or randomization:ti,ab,kw	568,926
#22	MeSH descriptor: [Randomized Controlled Trials as Topic] explode all trees	23,085
#23	#21 or #22	569,322
#24	drug*:ti,ab,kw or medicine*:ti,ab,kw	394,028
#25	MeSH descriptor: [Drug Therapy] explode all trees	138,470
#26	#24 or #25	446,327
#27	#20 and #23 and #26	56,186
	Limit to Cochrane reviews only	341

Part 3: Search for previous meta-epidemiological studies (MEDLINE via PubMed)

Previous meta-epidemiological studies were identified from an umbrella Cochrane review of studies comparing treatment effects from RCTs and NRS (Anglemyer et al., 2014). The searches conducted by Anglemyer et al. (2014) were updated from December 2013 until October 2018 (search date: 8 October 2018) to identify additional meta-epidemiological studies published since the umbrella review was conducted. The search terms were modified to yield a manageable number of studies and focus on pharmacological interventions.

Table A 3: Search strategy for meta-epidemiological studies published since December 2013 (MEDLINE via PubMed)

Search #	**Search terms**	**Number of records**

#1	Cochrane Database Syst Rev [TA] OR search[tiab] OR meta-analysis[PT] OR MEDLINE[tiab] OR PubMed[tiab] OR (systematic*[tiab] AND review*[tiab])	438,486
#2	"Observation"[mh] OR "Cohort Studies"[mh] OR "Longitudinal Studies"[mh] OR "Retrospective Studies"[mh] OR "Prospective Studies"[mh] OR observational[tiab] OR cohort*[tiab] OR crosssectional[tiab] OR crossectional[tiab] OR cross-sectional[tiab] OR cross sectional[tiab] OR longitudinal[tiab] OR causal inference*[tw] OR causality[tw] OR "instrumental variable"[tw] OR "structural model"[tw] OR practice-based[tw] OR propensity score*[tw] OR natural experiment*[tw] OR case-control[tw] OR before-after[tw] OR pre-post[tw] OR case-cohort[tw] OR case-crossover[tw] OR serial[tiab] OR nonexperimental[tiab] OR non-experimental[tiab] OR "nonrandomized"[tiab] OR "nonrandomised"[tiab] OR "non-randomised"[tiab] OR "nonrandomised"[tiab] OR "study designs"[tiab] OR "newcastle ottawa"[tiab] OR overestimat*[tiab] OR over-estimat*[tiab] OR bias[tiab] OR "are needed"[tiab] OR (evidence[tiab] AND quality[tiab])	2,873,431
#3	compara*[tiab] OR comparison*[tiab] OR contrast*[tiab] OR similar*[tiab] OR consistent*[tiab] OR inconsistent*[tiab] OR dissimilar*[tiab] OR differen*[tiab] OR concordan*[tiab] OR discordan*[tiab] OR heterogene*[tiab] OR "Research Design"[mh]	8,499,330
#4	(pharmacologic therapy[MeSH Terms]) OR drug*[Text Word] OR medicine*[Text Word])	6,339,992
#5	((((#1) AND #2) AND #3) AND #4)	21,751
#6	Limit #5 to only humans	19,761
#7	Limit #6 to December 2013-now	8,305

References for Appendix 1:

Anglemyer A, Horvath HT, Bero L. Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials. Cochrane Database Syst Rev. 2014(4): MR000034. Epub 2014/05/02.

Ioannidis JP, Haidich AB, Pappa M, Pantazis N, Kokori SI, Tektonidou MG, et al. Comparison of evidence of treatment effects in randomized and nonrandomized studies. JAMA. 2001;286(7): 821-30. Epub 2001/08/22.

Appendix 3: Data extracted for chapter 3

Variable name		Description
Meta analysis-level information	**Source meta-analysis ID**	Unique ID for each source meta-analysis (ID is identical for meta-epidemiological studies contributing more than 1 topic).
	Source meta-analysis	Name and publication year of source meta-analysis.
	Topic ("drug" and "outcome" in "patients")	Brief topic description through combination of the name of the drug, outcome, and participants.
	Therapeutic area (ATC code first level)	First-level ATC code under which the drug is categorized.
	Participants	Brief description of participants, as reported in source meta-analysis.
	Intervention	Name of the pharmacological intervention.
	Comparator	Name of the comparator.
	Outcome	Outcome of interest, as reported in source meta-analysis.
	Beneficial vs. detrimental outcome	"Beneficial" for outcomes where an increase in the number of participants experiencing an event, or an increase in the measured scale, is desired. "Detrimental" for outcomes where an increase in the number of participants experiencing an event, or an increase in the measured scale, is not desired.
	Do authors report pooled estimate within text?	Indication of whether single pooled effect estimate (obtained from both RCT and NRS in the same meta-analysis) is reported in the abstract, main body of text, or only included in a figure.
	Studies included in meta-analysis	Total number of studies with information contributing to the pooled effect.
Study-level information	**Study name**	Name of study, as reported in source meta-analysis.
	Year	Year of study, as reported in source meta-analysis.
	PMID	Unique PubMed ID for study, if available.
	Other	Unique ID for study, if not indexed in PubMed.
	Study type	RCT or non-randomized study, as reported in the source meta-analysis. ○ RCT: study is described as RCT in source meta-analysis.

		○ NRS: study is described as non- randomized study in source meta-analysis. For Cochrane reviews: study does not use a random sequence generation.
	Study type details	Study type categorization, based on information provided in source meta-analysis. ○ Experimental: applies to NRS only; investigators assign participants to intervention or control arm (e.g. in a non-randomized controlled trial) ○ Observational: applies to NRS only; no experimental set-up of the study ○ Individual RCT: applies to RCT only; randomization done at the individual level ○ Cluster RCT: applies to RCTs only; randomization done at the group/practice/hospital/regional level ○ Unclear: insufficient information to categorize study
	NRS type	Non- randomized study type categorization, based on information provided in source meta-analysis. ○ Quasi- randomized: a quasi-random allocation sequence was used, such as date of birth; social insurance number; odd vs. even date; etc. ○ Non- randomized: participants were allocated to intervention and control arm in a clinical trial based on investigator's or physician's preference (no measures taken to ensure random or quasi-random allocation). ○ Retrospective Cohort: participants who have received an intervention are retrospectively compared to those who have not received that intervention. ○ Prospective Cohort: participants who receive an intervention are compared to those who do not receive that intervention, and the study is planned prospectively. ○ Case-Control: retrospective studies where study participants are identified through their outcome. ○ Controlled before-and-after ○ Cross-sectional ○ Retrospective analysis of RCT: only applies when randomization was not preserved ○ Unclear/other ○ N/A: only applies to RCTs
	NRS data source	Non- randomized study data source categorization, based on information provided in source meta-analysis. ○ Prospectively collected data for study: data collected for the purpose of the study

		○ Disease registry ○ Product registry ○ Hospital case records ○ Primary or secondary care administrative database ○ Claims database ○ Other/unclear data sources ○ N/A: only applies to RCTs
	NRS data source: other	For non- randomized studies with "other" data source, the type of data source is described.
	NRS analytical method	For non- randomized studies only, the type of analytical method was extracted, as reported in the source meta-analysis, or inferred based on information provided in the source meta-analysis.
	NRS type of control	For NRS only: which type of control was used. Information either as reported in source meta-analysis or, where possible, inferred from description of study in source meta-analysis. Options include: ○ Concurrent: Participants in intervention and control arm are recruited from the same pool of patients, so that they could conceivably have received either the novel or the standard (control) treatment. ○ Historical: "In historical cohort studies, the treated cohort is compared with an untreated cohort who did not receive the intervention in a previous period, i.e. the individuals are not studied concurrently." (NICE DSU Technical Support Document 17) ○ Self-control: the control group consists of the same individuals as those receiving the experimental intervention. E.g. wound dressing (half of the wound treated with new intervention, other half with standard treatment), or comparisons within individuals over time (observation without intervention, followed by observation with intervention). ○ Unclear: insufficient information in source meta-analysis to categorize type of control. ○ N/A: only applies to RCTs
	Study sites	○ Single-center: includes studies where information was provided on the exact center/hospital where the study took place, or reference is made in the source meta-analysis to the study being "single-center" ○ Multi-center: includes studies where more than one country/city, or more than one center, are listed, or

		reference is made in the source meta-analysis to the study being "multi-center" o Unclear: includes studies where reference is made to a single country or city, but no specific study site is mentioned. o Not reported: no information about study sites were available in the source meta-analysis.
	Comparator type	o Active o Placebo/No treatment: includes standard of care
	Outcome type	Outcomes were categorized into three types. In case of doubts, Wood et al. (*BMJ,* 2008) was used as key reference. For composite outcomes, the most subjective component was used to categorize the study as objective or subjective. o Mortality o Objective o Subjective
	Participant details	Details about study participants (inclusion criteria), as reported in the source meta-analysis. If no study-level information was provided, the definition for population as used in the meta-analysis was extracted.
	Intervention details (dosing, administration, treatment period)	Details about the intervention, as reported in the source meta-analysis. If no study-level information was provided, the definition for intervention as used in the meta-analysis was extracted.
	Risk of bias tool	Name of the tool that was used to assess risk of bias, as reported in the source meta-analysis.
	Risk of bias score	Overall score or overall judgement for the risk of bias tool used, as reported in the source meta-analysis.
	Risk of Bias categorization	Standardization of risk of bias scores into "high", "moderate" and "low" risk of bias.
	N total sample size	Number of participants in intervention and control arms of the study. Function of numbers of participants in each arm.
	N intervention arm	Number of participants in the intervention arm of the study, as reported in the source meta-analysis.
	N control arm	Number of participants in the control arm of the study, as reported in the source meta-analysis.
	Effect measure	Effect measure for pooled effect estimate, as reported in the source meta-analysis. Including: RR=risk ratio; OR=odds ratio; HR=hazard ratio; RD=risk difference; MD=mean difference; SMD=standardized mean difference
	Coding of effect measure	Information on how the relationship between intervention and control arm was presented in the source meta-analysis:

		○ Intervention vs. control: For relative effect measures (RR, OR, HR): events in intervention arm / events in control arm For absolute effect measures (RD, MD, SMD): effect in intervention arm – effect in control arm ○ Control vs. intervention: For relative effect measures (RR, OR, HR): events in control arm / events in intervention arm For absolute effect measures (RD, MD, SMD): effect in control arm – effect in intervention arm
	Effect estimate as reported	Effect estimate, as reported in the source meta-analysis.
	Measure of variance	Indicates which measure of variance was reported in the source meta-analysis (including standard error, 95% confidence interval).
	SE as reported	Standard error of effect estimate, as reported in the source meta-analysis (N/A if not reported).
	Lower 95% CI as reported	Lower bound of 95% confidence interval of effect estimate, as reported in the source meta-analysis.
	Upper 95% CI as reported	Upper bound of 95% confidence interval of effect estimate, as reported in the source meta-analysis.
	p-value as reported	p-value for effect estimate, if reported in the source meta-analysis (N/A if not reported).

Appendix 4: Characteristics of meta-analyses included in chapter 3

Source meta-analysis	Comparator	Outcome type	Therapeutic area by WHO ATC code first level categorization	Risk of bias across NRSs in a meta-analysis	Risk of bias across RCTs in a meta-analysis	Median publication year of the studies included in the meta-analyses	Matching quality of RCTs and NRSs in meta-analysis	Timing of evidence generation	Cochrane review	Meta-analysis published in a top journal
Abolhassani 2013	Active	Other objective outcome	Anti-infective for systemic use	Moderate median risk of bias	Low median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	No
Afolabi 2012	Active	Other objective outcome	Nervous system	High median risk of bias	Moderate median risk of bias	Before 2000	Moderate (score of 7-9 of 12)	NRS published before first RCT	Yes	Yes
Agarwal 2017	Placebo or no treatment	Subjective outcome	Blood and blood forming organs	Low median risk of bias	No risk of bias information	2010 and later	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	No
Agarwal 2018	Active	Subjective outcome	Blood and blood forming organs	Low median risk of bias	Moderate median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	No

Alfirevic 2009	Placebo or no treatment	Other objective outcome	Systemic hormonal preparations	High median risk of bias	High median risk of bias	Before 2000	Moderate (score of 7-9 of 12)	First RCT published before NRS	Yes	Yes
Allen 2010	Placebo or no treatment	Subjective outcome	Alimentary tract and metabolism	High median risk of bias	High median risk of bias	2000-2009	Low (score of 4-6 of 12)	First NRS and first RCT published in the same year	Yes	Yes
Ampuero 2016	Placebo or no treatment	Other objective outcome	Antineoplastic and immuno-modulating agents	No risk of bias information	No risk of bias information	2010 and later	High (score of 10-12 of 12)	First NRS and first RCT published in the same year	No	No
An 2017	Placebo or no treatment	Other objective outcome	Cardiovascular system	Low median risk of bias	Moderate median risk of bias	2000-2009	Low (score of 4-6 of 12)	NRS published before first RCT	No	No
Andia 2014	Active	Subjective outcome	Blood and blood forming organs	Moderate median risk of bias	High median risk of bias	2010 and later	High (score of 10-12 of 12)	First NRS and first RCT published in the same year	No	No
Antoniou 2014	Placebo or no treatment	Mortality	Cardiovascular system	Low median risk of bias	Moderate median risk of bias	2000-2009	Low (score of 4-6 of 12)	First RCT published before NRS	No	No

Araujo 2015	Placebo or no treatment	Mortality	Antineoplastic and immuno-modulating agents	Low median risk of bias	Moderate median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	No
Arnaud 2014	Placebo or no treatment	Subjective outcome	Blood and blood forming organs	No risk of bias information	No risk of bias information	2000-2009	High (score of 10-12 of 12)	NRS published before first RCT	No	No
Austin 2015	Placebo or no treatment	Other objective outcome	Alimentary tract and metabolism	High median risk of bias	Moderate median risk of bias	Before 2000	Moderate (score of 7-9 of 12)	NRS published before first RCT	Yes	Yes
Ayoub 2016	Placebo or no treatment	Mortality	Blood and blood forming organs	Low median risk of bias	Low median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	No
Bai 2015	Placebo or no treatment	Subjective outcome	Dermatologicals	High median risk of bias	Low median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	No
Bakhshehian 2015	Placebo or no treatment	Other objective outcome	Anti-infective for systemic use	Moderate median risk of bias	Low median risk of bias	2010 and later	High (score of 10-12 of 12)	NRS published before first RCT	No	No
Baldinger 2012	Placebo or no treatment	Subjective outcome	Alimentary tract and metabolism	High median risk of bias	High median risk of bias	Before 2000	Moderate (score of 7-9 of 12)	First NRS and first RCT published in	Yes	Yes

								the same year		
Ballinger 2014	Active	Other objective outcome	Anti-infective for systemic use	High median risk of bias	High median risk of bias	Before 2000	Moderate (score of 7-9 of 12)	NRS published before first RCT	Yes	Yes
Bang 2015	Placebo or no treatment	Other objective outcome	Musculo-skeletal system	Moderate median risk of bias	Low median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	No
Barkat 2017	Placebo or no treatment	Subjective outcome	Blood and blood forming organs	Low median risk of bias	Moderate median risk of bias	2010 and later	High (score of 10-12 of 12)	First RCT published before NRS	No	No
Bellemain-Appaix	Placebo or no treatment	Mortality	Blood and blood forming organs	Low median risk of bias	High median risk of bias	2000-2009	Low (score of 4-6 of 12)	First NRS and first RCT published in the same year	No	Yes
Benjo 2016	Placebo or no treatment	Other objective outcome	Cardiovascular system	Low median risk of bias	Low median risk of bias	2000-2009	Low (score of 4-6 of 12)	NRS published before first RCT	No	No
Bhangu 2014	Placebo or no treatment	Other objective outcome	Musculo-skeletal system	Low median risk of bias	Low median risk of bias	2000-2009	Low (score of 4-6 of 12)	First RCT published before NRS	No	No

Bloom 2012	Active	Subjective outcome	Systemic hormonal preparations	High median risk of bias	High median risk of bias	2000-2009	High (score of 10-12 of 12)	First NRS and first RCT published in the same year	Yes	Yes
Bonet 2017	Placebo or no treatment	Subjective outcome	Anti-infective for systemic use	High median risk of bias	High median risk of bias	Before 2000	Low (score of 4-6 of 12)	NRS published before first RCT	Yes	Yes
Bosanquet 2015	Placebo or no treatment	Other objective outcome	Nervous system	High median risk of bias	Moderate median risk of bias	Before 2000	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	No
Bossard 2017	Placebo or no treatment	Other objective outcome	Blood and blood forming organs	Low median risk of bias	High median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	No
Boyle 2012	Placebo or no treatment	Subjective outcome	Anti-infective for systemic use	High median risk of bias	High median risk of bias	2000-2009	Low (score of 4-6 of 12)	First RCT published before NRS	Yes	Yes
Branger 2016	Active	Other objective outcome	Antineoplastic and immuno-modulating agents	High median risk of bias	Moderate median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	First NRS and first RCT published in the same year	No	Yes

Brennan 2012	Placebo or no treatment	Subjective outcome	Dermatologicals	High median risk of bias	Moderate median risk of bias	Before 2000	Low (score of 4-6 of 12)	First RCT published before NRS	Yes	Yes
Brito 2017	Placebo or no treatment	Subjective outcome	Antiparasitic products	Moderate median risk of bias	High median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	Yes
Brogly 2014	Placebo or no treatment	Subjective outcome	Nervous system	No risk of bias information	No risk of bias information	2010 and later	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	No
Brustia 2016	Both active and placebo-controlled studies	Other objective outcome	Blood and blood forming organs	Moderate median risk of bias	High median risk of bias	2010 and later	Low (score of 4-6 of 12)	First RCT published before NRS	No	No
Budden 2014	Active	Other objective outcome	Systemic hormonal preparations	High median risk of bias	High median risk of bias	Before 2000	High (score of 10-12 of 12)	First RCT published before NRS	Yes	Yes
Caldwell 2016	Placebo or no treatment	Subjective outcome	Nervous system	High median risk of bias	High median risk of bias	Before 2000	Moderate (score of 7-9 of 12)	NRS published before first RCT	Yes	Yes
Campbell 2017	Placebo or no treatment	Other objective outcome	Anti-infective for systemic use	High median risk of bias	Moderate median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	NRS published before first RCT	Yes	Yes

Carneiro 2015	Placebo or no treatment	Mortality	Antineoplastic and immuno-modulating agents	No risk of bias information	No risk of bias information	2000-2009	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	No
Chai-Adisaksopha 2016	Active	Mortality	Blood and blood forming organs	Moderate median risk of bias	High median risk of bias	2010 and later	High (score of 10-12 of 12)	First RCT published before NRS	No	No
Chalhoub 2017	Placebo or no treatment	Subjective outcome	Antineoplastic and immuno-modulating agents	Low median risk of bias	No risk of bias information	2010 and later	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	No
Chao 2017	Placebo or no treatment	Subjective outcome	Genito-urinary system and sex hormones	Low median risk of bias	Moderate median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	First NRS and first RCT published in the same year	No	No
Chen 2013	Placebo or no treatment	Other objective outcome	Anti-infective for systemic use	No risk of bias information	No risk of bias information	2000-2009	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	Yes
Chen 2015	Placebo or no treatment	Subjective outcome	Blood and blood forming organs	No risk of bias information	No risk of bias information	2010 and later	High (score of 10-12 of 12)	NRS published before first RCT	No	No

Chen 2015	Placebo or no treatment	Other objective outcome	Anti-infective for systemic use	High median risk of bias	Moderate median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	No
Chen 2017	Placebo or no treatment	Other objective outcome	Antineoplastic and immuno-modulating agents	Low median risk of bias	High median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	No
Cheng 2015	Placebo or no treatment	Subjective outcome	Sensory organs	No risk of bias information	Low median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	No
Chowdhury 2018	Placebo or no treatment	Mortality	Dermatologicals	High median risk of bias	Moderate median risk of bias	Before 2000	High (score of 10-12 of 12)	First RCT published before NRS	No	No
Chrcanovic 2014	Placebo or no treatment	Other objective outcome	Anti-infective for systemic use	High median risk of bias	Low median risk of bias	2000-2009	Low (score of 4-6 of 12)	First RCT published before NRS	No	No
Clifton 2014	Placebo or no treatment	Subjective outcome	Alimentary tract and metabolism	No risk of bias information	No risk of bias information	2010 and later	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	No
Coppola 2015	Placebo or no treatment	Other objective outcome	Blood and blood forming organs	High median risk of bias	Moderate median risk of bias	Before 2000	High (score of 10-12 of 12)	NRS published before first RCT	Yes	Yes
Costi 2014	Placebo or no treatment	Subjective outcome	Nervous system	High median risk of bias	Moderate median risk of bias	Before 2000	High (score of 10-12 of 12)	First RCT published before NRS	Yes	Yes

Coussement 2018	Placebo or no treatment	Other objective outcome	Anti-infective for systemic use	High median risk of bias	High median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	NRS published before first RCT	Yes	Yes
Critchley 2014	Placebo or no treatment	Mortality	Systemic hormonal preparations	High median risk of bias	Moderate median risk of bias	Before 2000	Low (score of 4-6 of 12)	First RCT published before NRS	Yes	Yes
Cui 2014	Both active and placebo-controlled studies	Other objective outcome	Blood and blood forming organs	No risk of bias information	No risk of bias information	2000-2009	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	No
Dahal 2015	Placebo or no treatment	Other objective outcome	Cardiovascular system	No risk of bias information	Low median risk of bias	2010 and later	High (score of 10-12 of 12)	First RCT published before NRS	No	No
David 2017	Active	Subjective outcome	Systemic hormonal preparations	High median risk of bias	High median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	First RCT published before NRS	Yes	Yes
de Frutos 2015	Placebo or no treatment	Other objective outcome	Cardiovascular system	High median risk of bias	Moderate median risk of bias	Before 2000	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	No
Desiderio 2017	Placebo or no treatment	Mortality	Antineoplastic and immuno-modulating agents	Low median risk of bias	Moderate median risk of bias	2000-2009	Low (score of 4-6 of 12)	NRS published before first RCT	No	No

Di 2015	Active	Other objective outcome	Anti-infective for systemic use	Low median risk of bias	Low median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	No
Dong 2017	Active	Other objective outcome	Alimentary tract and metabolism	Low median risk of bias	Moderate median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	First NRS and first RCT published in the same year	No	No
Edmonds 2012	Placebo or no treatment	Other objective outcome	Respiratory system	High median risk of bias	Moderate median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	First RCT published before NRS	Yes	Yes
El Sayed 2018	Active	Other objective outcome	Anti-infective for systemic use	High median risk of bias	High median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	First RCT published before NRS	Yes	Yes
Elgendy 2017	Active	Subjective outcome	Blood and blood forming organs	Low median risk of bias	Low median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	No
Engelen 2018	Placebo or no treatment	Subjective outcome	Blood and blood forming organs	High median risk of bias	Moderate median risk of bias	Before 2000	High (score of 10-12 of 12)	First RCT published before NRS	Yes	Yes
Engelman 2010	Placebo or no treatment	Other objective outcome	Systemic hormonal preparations	High median risk of bias	High median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	No

Estcourt 2015	Placebo or no treatment	Mortality	Antineoplastic and immuno-modulating agents	High median risk of bias	High median risk of bias	Before 2000	Moderate (score of 7-9 of 12)	First RCT published before NRS	Yes	Yes
Facciorusso 2018	Both active and placebo-controlled studies	Mortality	Cardiovascular system	Low median risk of bias	High median risk of bias	2010 and later	Low (score of 4-6 of 12)	First NRS and first RCT published in the same year	No	No
Falagas 2013	Active	Other objective outcome	Anti-infective for systemic use	No risk of bias information	No risk of bias information	2000-2009	High (score of 10-12 of 12)	First RCT published before NRS	No	Yes
Feng 2015	Both active and placebo-controlled studies	Other objective outcome	Alimentary tract and metabolism	No risk of bias information	No risk of bias information	2010 and later	High (score of 10-12 of 12)	First RCT published before NRS	No	No
Ferrer 2016	Placebo or no treatment	Other objective outcome	Anti-infective for systemic use	No risk of bias information	No risk of bias information	2000-2009	Low (score of 4-6 of 12)	First NRS and first RCT published in the same year	No	No
Filippini 2017	Placebo or no treatment	Subjective outcome	Antineoplastic and immuno-modulating agents	High median risk of bias	High median risk of bias	2010 and later	High (score of 10-12 of 12)	First RCT published before NRS	Yes	Yes

Fukuta 2017	Placebo or no treatment	Mortality	Cardiovascular system	Low median risk of bias	Low median risk of bias	2000-2009	Low (score of 4-6 of 12)	First RCT published before NRS	No	No
Fung 2015	Active	Other objective outcome	Anti-infective for systemic use	Low median risk of bias	Moderate median risk of bias	2010 and later	Low (score of 4-6 of 12)	First RCT published before NRS	No	Yes
Furtado 2014	Placebo or no treatment	Subjective outcome	Various	No risk of bias information	Low median risk of bias	2000-2009	High (score of 10-12 of 12)	First RCT published before NRS	No	No
Galappaththy 2013	Placebo or no treatment	Other objective outcome	Antiparasitic products	High median risk of bias	Moderate median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	First RCT published before NRS	Yes	Yes
Gandhi 2015	Placebo or no treatment	Subjective outcome	Blood and blood forming organs	Moderate median risk of bias	High median risk of bias	2010 and later	High (score of 10-12 of 12)	First RCT published before NRS	No	No
Gausden 2017	Placebo or no treatment	Other objective outcome	Blood and blood forming organs	No risk of bias information	No risk of bias information	2010 and later	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	No
Gharaibeh 2013	Placebo or no treatment	Subjective outcome	Blood and blood forming organs	High median risk of bias	High median risk of bias	Before 2000	High (score of 10-12 of 12)	First RCT published before NRS	Yes	Yes
Gillespie 2010	Placebo or no treatment	Subjective outcome	Anti-infective for systemic use	High median risk of bias	High median risk of bias	Before 2000	Moderate (score of 7-9 of 12)	First RCT published before NRS	Yes	Yes

Gong 2017	Placebo or no treatment	Subjective outcome	Sensory organs	Low median risk of bias	Moderate median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	Yes
Gonzalez 2018	Active	Other objective outcome	Antiparasitic products	High median risk of bias	Low median risk of bias	2010 and later	High (score of 10-12 of 12)	NRS published before first RCT	Yes	Yes
Grabein 2017	Active	Subjective outcome	Anti-infective for systemic use	High median risk of bias	Moderate median risk of bias	Before 2000	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	No
Graves 2010	Placebo or no treatment	Other objective outcome	Anti-infective for systemic use	No risk of bias information	No risk of bias information	Before 2000	High (score of 10-12 of 12)	NRS published before first RCT	Yes	Yes
Graves 2018	Placebo or no treatment	Other objective outcome	Antiparasitic products	High median risk of bias	High median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	First RCT published before NRS	Yes	Yes
Gray 2016	Placebo or no treatment	Mortality	Cardiovascular system	Low median risk of bias	Moderate median risk of bias	2010 and later	Low (score of 4-6 of 12)	NRS published before first RCT	No	No
Guerra 2017	Both active and placebo-controlled studies	Subjective outcome	Cardiovascular system	High median risk of bias	Low median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	No

Gunter 2017	Both active and placebo-controlled studies	Subjective outcome	Musculo-skeletal system	No risk of bias information	No risk of bias information	2000-2009	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	No
Haas 2015	Active	Mortality	Dermatologicals	High median risk of bias	High median risk of bias	Before 2000	High (score of 10-12 of 12)	First RCT published before NRS	Yes	Yes
Han 2016	Active	Other objective outcome	Blood and blood forming organs	No risk of bias information	No risk of bias information	2010 and later	High (score of 10-12 of 12)	First NRS and first RCT published in the same year	No	No
Han 2017	Active	Other objective outcome	Anti-infective for systemic use	Low median risk of bias	Moderate median risk of bias	2010 and later	High (score of 10-12 of 12)	First RCT published before NRS	No	No
Hannah 2016	Active	Other objective outcome	Dermatologicals	Moderate median risk of bias	High median risk of bias	2010 and later	High (score of 10-12 of 12)	NRS published before first RCT	No	No
Hao 2016	Active	Other objective outcome	Anti-infective for systemic use	Moderate median risk of bias	Moderate median risk of bias	2000-2009	High (score of 10-12 of 12)	NRS published before first RCT	No	No
Harnoss 2017	Placebo or no treatment	Subjective outcome	Anti-infective for systemic use	High median risk of bias	High median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	No

Haroon 2014	Placebo or no treatment	Other objective outcome	Antineoplastic and immuno-modulating agents	Moderate median risk of bias	No risk of bias information	2010 and later	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	No
Hass 2018	Placebo or no treatment	Subjective outcome	Genito-urinary system and sex hormones	High median risk of bias	High median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	First RCT published before NRS	Yes	Yes
He 2013	Active	Other objective outcome	Antineoplastic and immuno-modulating agents	Low median risk of bias	Low median risk of bias	2010 and later	High (score of 10-12 of 12)	First RCT published before NRS	No	Yes
He 2015	Active	Subjective outcome	Cardiovascular system	Low median risk of bias	Low median risk of bias	2000-2009	Low (score of 4-6 of 12)	First RCT published before NRS	No	No
Heal 2017	Placebo or no treatment	Other objective outcome	Dermatologicals	High median risk of bias	Low median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	No
Hemkens 2016	Placebo or no treatment	Mortality	Cardiovascular system	Moderate median risk of bias	High median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	First NRS and first RCT published in the same year	No	Yes
Hemkens 2016	Placebo or no treatment	Mortality	Blood and blood forming organs	Moderate median risk of bias	High median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	NRS published	No	Yes

								before first RCT		
Hemkens 2016	Placebo or no treatment	Mortality	Blood and blood forming organs	Moderate median risk of bias	Moderate median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	Yes
Henderson-Smart 2010	Placebo or no treatment	Subjective outcome	Respiratory system	Moderate median risk of bias	High median risk of bias	Before 2000	Moderate (score of 7-9 of 12)	First NRS and first RCT published in the same year	Yes	Yes
Henssler 2016	Placebo or no treatment	Subjective outcome	Nervous system	High median risk of bias	High median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	No
Hernandez 2017	Active	Other objective outcome	Anti-infective for systemic use	Low median risk of bias	High median risk of bias	Before 2000	High (score of 10-12 of 12)	First NRS and first RCT published in the same year	No	No
Hodson 2013	Placebo or no treatment	Subjective outcome	Anti-infective for systemic use	High median risk of bias	High median risk of bias	Before 2000	Moderate (score of 7-9 of 12)	First RCT published before NRS	Yes	Yes
Hong 2014	Placebo or no treatment	Other objective outcome	Antineoplastic and immuno-modulating agents	Low median risk of bias	Low median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	No

Horbach 2016	Active	Other objective outcome	Antineoplastic and immuno-modulating agents	Moderate median risk of bias	Moderate median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	No
Horita 2016	Placebo or no treatment	Mortality	Anti-infective for systemic use	Low median risk of bias	High median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	No
Hu 2014	Placebo or no treatment	Other objective outcome	Musculo-skeletal system	No risk of bias information	No risk of bias information	2000-2009	Low (score of 4-6 of 12)	NRS published before first RCT	No	No
Hu 2015	Active	Mortality	Respiratory system	No risk of bias information	No risk of bias information	2000-2009	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	No
Hu 2016	Placebo or no treatment	Other objective outcome	Antineoplastic and immuno-modulating agents	No risk of bias information	No risk of bias information	2010 and later	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	No
Huang 2013	Placebo or no treatment	Mortality	Cardiovascular system	High median risk of bias	Low median risk of bias	2000-2009	High (score of 10-12 of 12)	First RCT published before NRS	No	No
Huang 2016	Placebo or no treatment	Subjective outcome	Blood and blood forming organs	Moderate median risk of bias	High median risk of bias	2010 and later	Low (score of 4-6 of 12)	NRS published	No	No

								before first RCT		
Huang 2017	Active	Subjective outcome	Anti-infective for systemic use	Low median risk of bias	Low median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	Yes
Huang 2017	Placebo or no treatment	Other objective outcome	Genito-urinary system and sex hormones	High median risk of bias	High median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	Yes
Hughes 2016	Placebo or no treatment	Subjective outcome	Systemic hormonal preparations	High median risk of bias	High median risk of bias	Before 2000	Low (score of 4-6 of 12)	First RCT published before NRS	Yes	Yes
Hunt 2010	Placebo or no treatment	Mortality	Blood and blood forming organs	High median risk of bias	Moderate median risk of bias	Before 2000	High (score of 10-12 of 12)	First RCT published before NRS	Yes	Yes
Hyun 2017	Placebo or no treatment	Other objective outcome	Anti-infective for systemic use	Low median risk of bias	High median risk of bias	2010 and later	High (score of 10-12 of 12)	NRS published before first RCT	No	No
Jain 2016	Active	Subjective outcome	Nervous system	High median risk of bias	High median risk of bias	2000-2009	High (score of 10-12 of 12)	First RCT published before NRS	No	No
Ji 2017	Placebo or no treatment	Other objective outcome	Antineoplastic and immuno-modulating agents	No risk of bias information	No risk of bias information	2010 and later	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	No

Jiang 2015	Placebo or no treatment	Other objective outcome	Genito-urinary system and sex hormones	Moderate median risk of bias	Moderate median risk of bias	2000-2009	High (score of 10-12 of 12)	First RCT published before NRS	No	No
Jiang 2016	Placebo or no treatment	Other objective outcome	Blood and blood forming organs	Low median risk of bias	Moderate median risk of bias	2010 and later	High (score of 10-12 of 12)	First NRS and first RCT published in the same year	No	No
Jian-Yu 2018	Placebo or no treatment	Mortality	Alimentary tract and metabolism	No risk of bias information	No risk of bias information	2010 and later	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	No
Johnston 2017	Active	Mortality	Anti-infective for systemic use	Moderate median risk of bias	Low median risk of bias	2010 and later	High (score of 10-12 of 12)	NRS published before first RCT	No	No
Kabra 2013	Placebo or no treatment	Subjective outcome	Anti-infective for systemic use	High median risk of bias	High median risk of bias	Before 2000	Low (score of 4-6 of 12)	NRS published before first RCT	Yes	Yes
Kalil 2009	Active	Other objective outcome	Anti-infective for systemic use	No risk of bias information	Moderate median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	Yes

Kamal 2017	Both active and placebo-controlled studies	Subjective outcome	Alimentary tract and metabolism	Moderate median risk of bias	High median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	No
Kamal 2017	Placebo or no treatment	Mortality	Cardiovascular system	Low median risk of bias	Low median risk of bias	2010 and later	Low (score of 4-6 of 12)	NRS published before first RCT	No	No
Kanbay 2014	Placebo or no treatment	Other objective outcome	Musculo-skeletal system	High median risk of bias	High median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	No
Kaplan 2016	Placebo or no treatment	Different types of outcomes	Antiparasitic products	Low median risk of bias	Moderate median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	No
Kenyon 2013	Active	Other objective outcome	Systemic hormonal preparations	High median risk of bias	High median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	First RCT published before NRS	Yes	Yes
Kessel 2015	Both active and placebo-controlled studies	Different types of outcomes	Anti-infective for systemic use	High median risk of bias	High median risk of bias	2010 and later	Low (score of 4-6 of 12)	NRS published before first RCT	No	No
Khan 2016	Active	Subjective outcome	Nervous system	Low median risk of bias	High median risk of bias	2010 and later	High (score of 10-12 of 12)	NRS published before first RCT	No	No

Khan 2017	Placebo or no treatment	Subjective outcome	Dermatologicals	Low median risk of bias	Low median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	No
Khan 2017	Placebo or no treatment	Mortality	Cardiovascular system	Moderate median risk of bias	Low median risk of bias	2000-2009	Low (score of 4-6 of 12)	NRS published before first RCT	No	No
Khoshbin 2013	Both active and placebo-controlled studies	Subjective outcome	Blood and blood forming organs	Low median risk of bias	Low median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	No
Kim 2016	Active	Other objective outcome	Respiratory system	Low median risk of bias	Moderate median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	Yes
Kirkland 2017	Placebo or no treatment	Other objective outcome	Respiratory system	High median risk of bias	No risk of bias information	2000-2009	Moderate (score of 7-9 of 12)	First RCT published before NRS	Yes	Yes
Kirsch 2017	Placebo or no treatment	Other objective outcome	Blood and blood forming organs	Low median risk of bias	Moderate median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	No
Kitsios 2015	Placebo or no treatment	Mortality	Blood and blood forming organs	No risk of bias information	Moderate median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	No

Kitsios 2015	Active	Mortality	Blood and blood forming organs	No risk of bias information	Moderate median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	No
Kitsios 2015	Both active and placebo-controlled studies	Mortality	Blood and blood forming organs	No risk of bias information	High median risk of bias	2000-2009	High (score of 10-12 of 12)	First RCT published before NRS	No	No
Kitsios 2015	Placebo or no treatment	Mortality	Cardiovascular system	No risk of bias information	High median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	No
Kitsios 2015	Placebo or no treatment	Mortality	Cardiovascular system	No risk of bias information	Low median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	No
Kitsios 2015	Placebo or no treatment	Mortality	Cardiovascular system	No risk of bias information	Low median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	No
Kitsios 2015	Placebo or no treatment	Mortality	Systemic hormonal preparations	No risk of bias information	Moderate median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	No
Kitsios 2015	Both active and placebo-controlled studies	Mortality	Blood and blood forming organs	No risk of bias information	Moderate median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	No

Kitsios 2015	Both active and placebo-controlled studies	Mortality	Blood and blood forming organs	No risk of bias information	Moderate median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	No
Klimo 2014	Placebo or no treatment	Other objective outcome	Anti-infective for systemic use	Moderate median risk of bias	Low median risk of bias	Before 2000	Low (score of 4-6 of 12)	NRS published before first RCT	No	No
Kovacs 2016	Placebo or no treatment	Other objective outcome	Antiparasitic products	Moderate median risk of bias	High median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	Yes
Kowalewski 2016	Active	Subjective outcome	Blood and blood forming organs	Low median risk of bias	Moderate median risk of bias	2000-2009	High (score of 10-12 of 12)	First RCT published before NRS	No	No
Krajewski	Active	Mortality	Blood and blood forming organs	Low median risk of bias	Moderate median risk of bias	2010 and later	High (score of 10-12 of 12)	NRS published before first RCT	No	No
Kroon 2015	Active	Subjective outcome	Musculo-skeletal system	High median risk of bias	Moderate median risk of bias	Before 2000	Moderate (score of 7-9 of 12)	First RCT published before NRS	Yes	Yes
Kuang 2017	Placebo or no treatment	Subjective outcome	Nervous system	Moderate median risk of bias	Low median risk of bias	2010 and later	High (score of 10-12 of 12)	First RCT published before NRS	No	No

Kwok 2013	Placebo or no treatment	Subjective outcome	Alimentary tract and metabolism	No risk of bias information	Low median risk of bias	2010 and later	High (score of 10-12 of 12)	NRS published before first RCT	No	No
Lee 2017	Active	Other objective outcome	Anti-infective for systemic use	Low median risk of bias	Moderate median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	No
Lee 2017	Active	Mortality	Anti-infective for systemic use	High median risk of bias	High median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	No
Leibovici-Weissman 2014	Placebo or no treatment	Subjective outcome	Anti-infective for systemic use	High median risk of bias	High median risk of bias	Before 2000	Low (score of 4-6 of 12)	NRS published before first RCT	Yes	Yes
Lemos 2014	Active	Subjective outcome	Antineoplastic and immuno-modulating agents	Low median risk of bias	High median risk of bias	2010 and later	High (score of 10-12 of 12)	First NRS and first RCT published in the same year	No	No
Leone 2016	Placebo or no treatment	Subjective outcome	Nervous system	High median risk of bias	High median risk of bias	Before 2000	Moderate (score of 7-9 of 12)	First RCT published before NRS	Yes	Yes
Lewis 2018	Both active and placebo-controlled studies	Mortality	Blood and blood forming organs	High median risk of bias	Moderate median risk of bias	2010 and later	Low (score of 4-6 of 12)	First RCT published before NRS	Yes	Yes

Li 2015	Placebo or no treatment	Other objective outcome	Antineoplastic and immuno-modulating agents	Moderate median risk of bias	Moderate median risk of bias	2010 and later	High (score of 10-12 of 12)	First RCT published before NRS	No	No
Li 2015	Placebo or no treatment	Other objective outcome	Sensory organs	Low median risk of bias	Low median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	Yes
Li 2016	Placebo or no treatment	Subjective outcome	Blood and blood forming organs	No risk of bias information	No risk of bias information	2010 and later	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	No
Li 2017	Placebo or no treatment	Subjective outcome	Musculo-skeletal system	Low median risk of bias	Low median risk of bias	2010 and later	High (score of 10-12 of 12)	First RCT published before NRS	No	No
Li 2018	Active	Subjective outcome	Antineoplastic and immuno-modulating agents	No risk of bias information	No risk of bias information	2010 and later	High (score of 10-12 of 12)	First RCT published before NRS	No	No
Liang 2014	Both active and placebo-controlled studies	Mortality	Antineoplastic and immuno-modulating agents	Moderate median risk of bias	Low median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	No
Liang 2017	Placebo or no treatment	Subjective outcome	Nervous system	Low median risk of bias	Moderate median risk of bias	2010 and later	High (score of 10-12 of 12)	First RCT published before NRS	No	No

Liet 2015	Placebo or no treatment	Other objective outcome	Various	High median risk of bias	Moderate median risk of bias	2000-2009	High (score of 10-12 of 12)	First RCT published before NRS	Yes	Yes
Lim 2015	Placebo or no treatment	Other objective outcome	Anti-infective for systemic use	Low median risk of bias	Moderate median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	No
Lim 2015	Placebo or no treatment	Other objective outcome	Cardiovascular system	No risk of bias information	Moderate median risk of bias	2010 and later	High (score of 10-12 of 12)	First RCT published before NRS	No	No
Lin 2015	Placebo or no treatment	Mortality	Antineoplastic and immuno-modulating agents	High median risk of bias	Low median risk of bias	2000-2009	Low (score of 4-6 of 12)	NRS published before first RCT	No	No
Liu 2013	Placebo or no treatment	Other objective outcome	Musculo-skeletal system	Low median risk of bias	Low median risk of bias	2010 and later	Low (score of 4-6 of 12)	NRS published before first RCT	No	No
Liu 2014	Placebo or no treatment	Subjective outcome	Anti-infective for systemic use	Moderate median risk of bias	Low median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	Yes
Liu 2015	Placebo or no treatment	Subjective outcome	Blood and blood forming organs	Moderate median risk of bias	High median risk of bias	2000-2009	High (score of 10-12 of 12)	First RCT published before NRS	No	Yes

Liu 2016	Placebo or no treatment	Other objective outcome	Dermatologicals	Low median risk of bias	Low median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	No
Liu 2017	Placebo or no treatment	Other objective outcome	Anti-infective for systemic use	Moderate median risk of bias	Moderate median risk of bias	2000-2009	High (score of 10-12 of 12)	First RCT published before NRS	No	Yes
Loomba 2015	Placebo or no treatment	Subjective outcome	Cardiovascular system	Low median risk of bias	No risk of bias information	2010 and later	High (score of 10-12 of 12)	NRS published before first RCT	No	No
Lu 2014	Placebo or no treatment	Other objective outcome	Anti-infective for systemic use	Low median risk of bias	Moderate median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	No
Luni 2018	Placebo or no treatment	Other objective outcome	Cardiovascular system	Low median risk of bias	High median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	No
Lussana 2014	Active	Subjective outcome	Blood and blood forming organs	No risk of bias information	No risk of bias information	2010 and later	High (score of 10-12 of 12)	First RCT published before NRS	No	No
Ma 2015	Placebo or no treatment	Subjective outcome	Blood and blood forming organs	High median risk of bias	High median risk of bias	2000-2009	Low (score of 4-6 of 12)	First RCT published before NRS	No	Yes

Mackeen 2011	Placebo or no treatment	Mortality	Genito-urinary system and sex hormones	High median risk of bias	High median risk of bias	Before 2000	Moderate (score of 7-9 of 12)	First RCT published before NRS	Yes	Yes
Mao 2015	Active	Other objective outcome	Systemic hormonal preparations	No risk of bias information	No risk of bias information	2000-2009	Low (score of 4-6 of 12)	First RCT published before NRS	No	No
Matthews 2016	Active	Other objective outcome	Systemic hormonal preparations	High median risk of bias	High median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	First RCT published before NRS	Yes	Yes
Mbeye 2014	Placebo or no treatment	Other objective outcome	Anti-infective for systemic use	No risk of bias information	No risk of bias information	2010 and later	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	No
Meduri 2016	Placebo or no treatment	Subjective outcome	Nervous system	Moderate median risk of bias	Moderate median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	No
Merlotti 2014	Placebo or no treatment	Other objective outcome	Cardiovascular system	High median risk of bias	High median risk of bias	2000-2009	Low (score of 4-6 of 12)	First RCT published before NRS	No	No
Mesfin 2016	Placebo or no treatment	Other objective outcome	Anti-infective for systemic use	Moderate median risk of bias	Moderate median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	No
Mesgarpour 2017	Placebo or no treatment	Subjective outcome	Blood and blood forming organs	Low median risk of bias	Moderate median risk of bias	2010 and later	Low (score of 4-6 of 12)	First RCT published before NRS	Yes	Yes

Miyake 2010	Placebo or no treatment	Other objective outcome	Antineoplastic and immuno-modulating agents	No risk of bias information	No risk of bias information	2000-2009	Low (score of 4-6 of 12)	NRS published before first RCT	No	No
Moraes 2014	Placebo or no treatment	Subjective outcome	Blood and blood forming organs	High median risk of bias	High median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	NRS published before first RCT	Yes	Yes
Muanda 2015	Placebo or no treatment	Other objective outcome	Antiparasitic products	High median risk of bias	Moderate median risk of bias	2000-2009	Low (score of 4-6 of 12)	First RCT published before NRS	No	Yes
Munnee 2016	Placebo or no treatment	Mortality	Alimentary tract and metabolism	No risk of bias information	Moderate median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	No
Muranushi 2015	Placebo or no treatment	Other objective outcome	Blood and blood forming organs	Low median risk of bias	Low median risk of bias	2010 and later	Low (score of 4-6 of 12)	NRS published before first RCT	No	No
Muranushi 2015	Placebo or no treatment	Other objective outcome	Blood and blood forming organs	High median risk of bias	Low median risk of bias	2010 and later	Low (score of 4-6 of 12)	NRS published before first RCT	No	No
Murphy 2016	Placebo or no treatment	Other objective outcome	Anti-infective for systemic use	High median risk of bias	Moderate median risk of bias	Before 2000	Low (score of 4-6 of 12)	NRS published	No	No

								before first RCT		
Muzii 2016	Active	Subjective outcome	Genito-urinary system and sex hormones	No risk of bias information	Moderate median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	No
Nairooz 2017	Placebo or no treatment	Subjective outcome	Blood and blood forming organs	Low median risk of bias	Low median risk of bias	2000-2009	High (score of 10-12 of 12)	First NRS and first RCT published in the same year	No	No
Neufeld 2016	Placebo or no treatment	Subjective outcome	Nervous system	High median risk of bias	High median risk of bias	2010 and later	Low (score of 4-6 of 12)	First RCT published before NRS	No	No
Niafar 2015	Placebo or no treatment	Other objective outcome	Alimentary tract and metabolism	Low median risk of bias	Low median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	No
Nie 2016	Placebo or no treatment	Mortality	Antineoplastic and immuno-modulating agents	Low median risk of bias	Moderate median risk of bias	2010 and later	High (score of 10-12 of 12)	NRS published before first RCT	No	No
O'Brien 2014	Placebo or no treatment	Subjective outcome	Genito-urinary system and sex hormones	No risk of bias information	No risk of bias information	2000-2009	Low (score of 4-6 of 12)	NRS published before first RCT	No	No

Ogunlesi 2015	Placebo or no treatment	Mortality	Systemic hormonal preparations	High median risk of bias	High median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	NRS published before first RCT	Yes	Yes
Ohlsson 2015	Placebo or no treatment	Mortality	Anti-infective for systemic use	High median risk of bias	High median risk of bias	Before 2000	Moderate (score of 7-9 of 12)	NRS published before first RCT	Yes	Yes
Okoli 2014	Placebo or no treatment	Subjective outcome	Anti-infective for systemic use	Moderate median risk of bias	No risk of bias information	2000-2009	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	Yes
Ortize-Orendain 2017	Placebo or no treatment	Subjective outcome	Nervous system	High median risk of bias	High median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	First RCT published before NRS	Yes	Yes
Ortiz-Salas 2016	Active	Subjective outcome	Anti-infective for systemic use	Low median risk of bias	Low median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	No
Osborn 2010	Active	Subjective outcome	Nervous system	High median risk of bias	High median risk of bias	Before 2000	High (score of 10-12 of 12)	First RCT published before NRS	Yes	Yes
Osborn 2010	Active	Subjective outcome	Nervous system	High median risk of bias	High median risk of bias	Before 2000	High (score of 10-12 of 12)	First RCT published before NRS	Yes	Yes
Paciaroni 2011	Both active and placebo-	Subjective outcome	Blood and blood forming organs	No risk of bias information	No risk of bias information	2000-2009	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	No

	controlled studies									
Pammi 2015	Placebo or no treatment	Mortality	Cardiovascular system	High median risk of bias	High median risk of bias	2000-2009	High (score of 10-12 of 12)	First RCT published before NRS	Yes	Yes
Pan 2015	Placebo or no treatment	Subjective outcome	Blood and blood forming organs	No risk of bias information	No risk of bias information	2010 and later	Low (score of 4-6 of 12)	NRS published before first RCT	No	No
Pan 2016	Both active and placebo-controlled studies	Other objective outcome	Nervous system	Moderate median risk of bias	Low median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	No
Pani 2014	Placebo or no treatment	Subjective outcome	Nervous system	High median risk of bias	High median risk of bias	2010 and later	Low (score of 4-6 of 12)	First RCT published before NRS	Yes	Yes
Paul 2014	Active	Mortality	Anti-infective for systemic use	High median risk of bias	High median risk of bias	Before 2000	Moderate (score of 7-9 of 12)	First RCT published before NRS	Yes	Yes
Paul 2016	Active	Mortality	Anti-infective for systemic use	High median risk of bias	Moderate median risk of bias	2010 and later	High (score of 10-12 of 12)	NRS published before first RCT	No	No
Paul 2016	Placebo or no treatment	Other objective outcome	Anti-infective for systemic use	Low median risk of bias	Moderate median risk of bias	2010 and later	Low (score of 4-6 of 12)	NRS published	No	Yes

								before first RCT		
Perez-Gaxiola 2018	Placebo or no treatment	Subjective outcome	Alimentary tract and metabolism	High median risk of bias	Moderate median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	First RCT published before NRS	Yes	Yes
Peters 2014	Both active and placebo-controlled studies	Subjective outcome	Cardiovascular system	No risk of bias information	No risk of bias information	2000-2009	Low (score of 4-6 of 12)	NRS published before first RCT	No	No
Prasad 2014	Placebo or no treatment	Subjective outcome	Nervous system	High median risk of bias	Low median risk of bias	Before 2000	Moderate (score of 7-9 of 12)	First RCT published before NRS	Yes	Yes
Price 2012	Active	Subjective outcome	Genito-urinary system and sex hormones	High median risk of bias	Moderate median risk of bias	Before 2000	High (score of 10-12 of 12)	NRS published before first RCT	Yes	Yes
Prijic 2014	Both active and placebo-controlled studies	Subjective outcome	Cardiovascular system	No risk of bias information	No risk of bias information	2000-2009	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	No
Prins 2015	Active	Mortality	Cardiovascular system	High median risk of bias	High median risk of bias	2000-2009	High (score of 10-12 of 12)	NRS published before first RCT	No	No

Proietti 2015	Active	Subjective outcome	Blood and blood forming organs	Low median risk of bias	Moderate median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	No
Prutsky 2013	Placebo or no treatment	Subjective outcome	Anti-infective for systemic use	Moderate median risk of bias	Moderate median risk of bias	Before 2000	Low (score of 4-6 of 12)	NRS published before first RCT	No	No
Puig 2016	Active	Other objective outcome	Alimentary tract and metabolism	Moderate median risk of bias	Moderate median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	No
Qin 2016	Placebo or no treatment	Mortality	Antineoplastic and immuno-modulating agents	Low median risk of bias	Moderate median risk of bias	2000-2009	Low (score of 4-6 of 12)	First RCT published before NRS	No	No
Qiu 2017	Placebo or no treatment	Other objective outcome	Antineoplastic and immuno-modulating agents	Moderate median risk of bias	Low median risk of bias	2000-2009	Low (score of 4-6 of 12)	NRS published before first RCT	No	No
Qiu 2018	Placebo or no treatment	Other objective outcome	Alimentary tract and metabolism	Low median risk of bias	Low median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	No
Radeva-Petrova 2014	Placebo or no treatment	Other objective outcome	Antiparasitic products	High median risk of bias	Moderate median risk of bias	Before 2000	Moderate (score of 7-9 of 12)	NRS published before first RCT	Yes	Yes

Rivero 2017	Placebo or no treatment	Subjective outcome	Respiratory system	Low median risk of bias	Low median risk of bias	2010 and later	High (score of 10-12 of 12)	NRS published before first RCT	No	No
Roberts 2018	Active	Other objective outcome	Anti-infective for systemic use	Moderate median risk of bias	Low median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	No
Rodriguez-Zuniga 2018	Placebo or no treatment	Subjective outcome	Anti-infective for systemic use	High median risk of bias	Moderate median risk of bias	2010 and later	High (score of 10-12 of 12)	NRS published before first RCT	No	No
Rojas-Villarraga 2014	Placebo or no treatment	Subjective outcome	Genito-urinary system and sex hormones	High median risk of bias	Moderate median risk of bias	2000-2009	Low (score of 4-6 of 12)	NRS published before first RCT	No	Yes
Rokkas 2014	Placebo or no treatment	Other objective outcome	Alimentary tract and metabolism	No risk of bias information	No risk of bias information	2010 and later	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	No
Rys 2018	Active	Other objective outcome	Alimentary tract and metabolism	No risk of bias information	No risk of bias information	2010 and later	High (score of 10-12 of 12)	NRS published before first RCT	No	No

Sahebkar 2017	Placebo or no treatment	Other objective outcome	Cardiovascular system	Moderate median risk of bias	Moderate median risk of bias	2010 and later	Low (score of 4-6 of 12)	NRS published before first RCT	No	No
Sahebkar 2017	Placebo or no treatment	Other objective outcome	Antineoplastic and immuno-modulating agents	Low median risk of bias	Moderate median risk of bias	Before 2000	Low (score of 4-6 of 12)	First RCT published before NRS	No	No
Salata 2018	Placebo or no treatment	Other objective outcome	Cardiovascular system	Low median risk of bias	Low median risk of bias	2010 and later	Low (score of 4-6 of 12)	NRS published before first RCT	No	No
Salvi 2017	Placebo or no treatment	Other objective outcome	Nervous system	Low median risk of bias	Low median risk of bias	2010 and later	Low (score of 4-6 of 12)	NRS published before first RCT	No	Yes
Sant'anna 2014	Active	Subjective outcome	Blood and blood forming organs	No risk of bias information	No risk of bias information	2010 and later	High (score of 10-12 of 12)	First NRS and first RCT published in the same year	No	No
Sardar 2014	Active	Subjective outcome	Blood and blood forming organs	Moderate median risk of bias	No risk of bias information	2010 and later	High (score of 10-12 of 12)	First NRS and first RCT published in the same year	No	No

Serpa Neto 2014	Placebo or no treatment	Other objective outcome	Blood and blood forming organs	No risk of bias information	Low median risk of bias	2010 and later	High (score of 10-12 of 12)	First RCT published before NRS	No	No
Seth 2014	Active	Subjective outcome	Musculo-skeletal system	High median risk of bias	High median risk of bias	2000-2009	High (score of 10-12 of 12)	First RCT published before NRS	Yes	Yes
Shang 2011	Active	Other objective outcome	Antineoplastic and immuno-modulating agents	High median risk of bias	High median risk of bias	Before 2000	High (score of 10-12 of 12)	NRS published before first RCT	Yes	Yes
Sharma 2014	Placebo or no treatment	Subjective outcome	Musculo-skeletal system	Moderate median risk of bias	No risk of bias information	2000-2009	Low (score of 4-6 of 12)	First RCT published before NRS	No	No
Shen 2016	Placebo or no treatment	Other objective outcome	Cardiovascular system	Low median risk of bias	Low median risk of bias	2010 and later	Low (score of 4-6 of 12)	NRS published before first RCT	No	No
Shi 2014	Placebo or no treatment	Other objective outcome	Cardiovascular system	Low median risk of bias	No risk of bias information	2010 and later	Low (score of 4-6 of 12)	NRS published before first RCT	No	Yes
Shi 2017	Placebo or no treatment	Subjective outcome	Blood and blood forming organs	High median risk of bias	High median risk of bias	2010 and later	High (score of 10-12 of 12)	NRS published before first RCT	No	No

Shim 2014	Placebo or no treatment	Other objective outcome	Genito-urinary system and sex hormones	Low median risk of bias	Low median risk of bias	Before 2000	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	No
Shin 2015	Active	Subjective outcome	Antineoplastic and immuno-modulating agents	Low median risk of bias	Low median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	First NRS and first RCT published in the same year	No	No
Sim 2010	Both active and placebo-controlled studies	Other objective outcome	Genito-urinary system and sex hormones	Moderate median risk of bias	High median risk of bias	2000-2009	High (score of 10-12 of 12)	First RCT published before NRS	No	No
Sinclair 2011	Placebo or no treatment	Other objective outcome	Anti-infective for systemic use	High median risk of bias	Moderate median risk of bias	Before 2000	Moderate (score of 7-9 of 12)	First RCT published before NRS	Yes	Yes
Singh 2017	Placebo or no treatment	Subjective outcome	Blood and blood forming organs	Moderate median risk of bias	Moderate median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	No
Singh 2017	Placebo or no treatment	Other objective outcome	Nervous system	High median risk of bias	High median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	No
Smaill 2014	Placebo or no treatment	Other objective outcome	Anti-infective for systemic use	High median risk of bias	Moderate median risk of bias	Before 2000	Moderate (score of 7-9 of 12)	First RCT published before NRS	Yes	Yes

Smit 2013	Placebo or no treatment	Subjective outcome	Nervous system	High median risk of bias	High median risk of bias	Before 2000	Moderate (score of 7-9 of 12)	First RCT published before NRS	Yes	Yes
Sole-Lleonart 2017	Active	Other objective outcome	Anti-infective for systemic use	Low median risk of bias	Low median risk of bias	2010 and later	High (score of 10-12 of 12)	NRS published before first RCT	No	No
Song 2017	Placebo or no treatment	Other objective outcome	Cardiovascular system	No risk of bias information	No risk of bias information	2010 and later	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	No
Sotiriadis 2015	Placebo or no treatment	Subjective outcome	Systemic hormonal preparations	High median risk of bias	Low median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	No
Squizzato 2010	Placebo or no treatment	Subjective outcome	Cardiovascular system	High median risk of bias	Low median risk of bias	2000-2009	Low (score of 4-6 of 12)	NRS published before first RCT	No	No
Stern 2014	Active	Subjective outcome	Anti-infective for systemic use	High median risk of bias	High median risk of bias	Before 2000	High (score of 10-12 of 12)	NRS published before first RCT	Yes	Yes
Strohmeier 2014	Active	Other objective outcome	Anti-infective for systemic use	High median risk of bias	High median risk of bias	2000-2009	High (score of 10-12 of 12)	First RCT published before NRS	Yes	Yes

Suthar 2015	Placebo or no treatment	Mortality	Anti-infective for systemic use	Low median risk of bias	Low median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	Yes
Talukdar 2015	Placebo or no treatment	Subjective outcome	Various	High median risk of bias	Moderate median risk of bias	2000-2009	High (score of 10-12 of 12)	First RCT published before NRS	No	No
Tang 2009	Placebo or no treatment	Mortality	Systemic hormonal preparations	No risk of bias information	No risk of bias information	2000-2009	Moderate (score of 7-9 of 12)	First NRS and first RCT published in the same year	No	No
Tang 2016	Placebo or no treatment	Other objective outcome	Cardiovascular system	Low median risk of bias	High median risk of bias	2010 and later	Low (score of 4-6 of 12)	NRS published before first RCT	No	No
Tarantini 2017	Active	Subjective outcome	Blood and blood forming organs	Low median risk of bias	Low median risk of bias	2010 and later	High (score of 10-12 of 12)	First RCT published before NRS	No	No
Taylor 2015	Placebo or no treatment	Other objective outcome	Anti-infective for systemic use	High median risk of bias	Moderate median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	First RCT published before NRS	Yes	Yes
Teng 2017	Placebo or no treatment	Other objective outcome	Cardiovascular system	Moderate median risk of bias	Moderate median risk of bias	2010 and later	High (score of 10-12 of 12)	NRS published before first RCT	No	No

Toews 2018	Active	Different types of outcomes	Alimentary tract and metabolism	High median risk of bias	High median risk of bias	Before 2000	High (score of 10-12 of 12)	NRS published before first RCT	Yes	Yes
Tran-Duy 2016	Placebo or no treatment	Subjective outcome	Alimentary tract and metabolism	Moderate median risk of bias	Moderate median risk of bias	2000-2009	Low (score of 4-6 of 12)	NRS published before first RCT	No	No
Tsai 2018	Active	Subjective outcome	Dermatologicals	Moderate median risk of bias	Low median risk of bias	2010 and later	High (score of 10-12 of 12)	First RCT published before NRS	No	No
Tsaousi 2016	Placebo or no treatment	Other objective outcome	Nervous system	High median risk of bias	Low median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	No
Tully 2016	Placebo or no treatment	Subjective outcome	Cardiovascular system	Low median risk of bias	Low median risk of bias	2000-2009	Low (score of 4-6 of 12)	First RCT published before NRS	No	No
Tunnicliffe 2018	Active	Other objective outcome	Antineoplastic and immuno-modulating agents	High median risk of bias	High median risk of bias	2000-2009	High (score of 10-12 of 12)	First RCT published before NRS	Yes	Yes
Turgeon 2015	Placebo or no treatment	Mortality	Blood and blood forming organs	Moderate median risk of bias	High median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	No

Ukaigwe 2017	Active	Subjective outcome	Blood and blood forming organs	Moderate median risk of bias	No risk of bias information	2010 and later	High (score of 10-12 of 12)	First NRS and first RCT published in the same year	No	No
van Herwaarden 2014	Active	Other objective outcome	Antineoplastic and immuno-modulating agents	High median risk of bias	High median risk of bias	2010 and later	High (score of 10-12 of 12)	First RCT published before NRS	Yes	Yes
Vardakas 2018	Placebo or no treatment	Mortality	Anti-infective for systemic use	No risk of bias information	Moderate median risk of bias	2010 and later	High (score of 10-12 of 12)	NRS published before first RCT	No	No
Vecchio 2015	Placebo or no treatment	Other objective outcome	Systemic hormonal preparations	High median risk of bias	High median risk of bias	Before 2000	Moderate (score of 7-9 of 12)	First RCT published before NRS	Yes	Yes
Vyas 2015	Active	Subjective outcome	Blood and blood forming organs	Low median risk of bias	Low median risk of bias	2010 and later	High (score of 10-12 of 12)	First RCT published before NRS	No	No
Wan 2018	Placebo or no treatment	Subjective outcome	Cardiovascular system	Moderate median risk of bias	Low median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	No
Wang 2014	Both active and placebo-	Subjective outcome	Cardiovascular system	No risk of bias information	No risk of bias information	2000-2009	High (score of 10-12 of 12)	NRS published	No	Yes

	controlled studies							before first RCT		
Wang 2015	Placebo or no treatment	Subjective outcome	Blood and blood forming organs	Moderate median risk of bias	Low median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	First NRS and first RCT published in the same year	No	No
Wang 2016	Placebo or no treatment	Subjective outcome	Musculo-skeletal system	Low median risk of bias	Moderate median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	No
Wang 2016	Placebo or no treatment	Other objective outcome	Anti-infective for systemic use	High median risk of bias	Moderate median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	No
Wang 2017	Both active and placebo-controlled studies	Other objective outcome	Alimentary tract and metabolism	No risk of bias information	High median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	No
Wang 2017	Placebo or no treatment	Other objective outcome	Antiparasitic products	Low median risk of bias	Low median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	Yes
Wang 2017	Placebo or no treatment	Other objective outcome	Systemic hormonal preparations	Moderate median risk of bias	Low median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	No

Wang 2018	Placebo or no treatment	Other objective outcome	Blood and blood forming organs	Low median risk of bias	Moderate median risk of bias	2010 and later	Low (score of 4-6 of 12)	First NRS and first RCT published in the same year	No	No
Watti 2017	Active	Mortality	Blood and blood forming organs	Low median risk of bias	No risk of bias information	2010 and later	High (score of 10-12 of 12)	First RCT published before NRS	No	No
Westhoff 2013	Both active and placebo-controlled studies	Subjective outcome	Systemic hormonal preparations	High median risk of bias	High median risk of bias	Before 2000	High (score of 10-12 of 12)	First RCT published before NRS	Yes	Yes
Whiting 2017	Active	Subjective outcome	Cardiovascular system	High median risk of bias	Moderate median risk of bias	2010 and later	High (score of 10-12 of 12)	First RCT published before NRS	No	Yes
Widmer 2015	Active	Other objective outcome	Anti-infective for systemic use	High median risk of bias	High median risk of bias	Before 2000	High (score of 10-12 of 12)	NRS published before first RCT	Yes	Yes
Wilhelmus 2015	Placebo or no treatment	Other objective outcome	Sensory organs	High median risk of bias	Moderate median risk of bias	Before 2000	Moderate (score of 7-9 of 12)	First NRS and first RCT published in the same year	Yes	Yes

Wiysonge 2017	Placebo or no treatment	Other objective outcome	Systemic hormonal preparations	High median risk of bias	Moderate median risk of bias	2000-2009	Low (score of 4-6 of 12)	NRS published before first RCT	Yes	Yes
Wu 2015	Both active and placebo-controlled studies	Other objective outcome	Alimentary tract and metabolism	Low median risk of bias	High median risk of bias	2000-2009	Low (score of 4-6 of 12)	First RCT published before NRS	No	No
Wu 2015	Placebo or no treatment	Other objective outcome	Cardiovascular system	Moderate median risk of bias	Moderate median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	Yes
Xia 2015	Placebo or no treatment	Mortality	Anti-infective for systemic use	Low median risk of bias	Low median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	No
Xia 2015	Placebo or no treatment	Subjective outcome	Antineoplastic and immuno-modulating agents	Low median risk of bias	High median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	No
Xie 2017	Placebo or no treatment	Other objective outcome	Antineoplastic and immuno-modulating agents	Moderate median risk of bias	High median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	No
Xing 2016	Placebo or no treatment	Other objective outcome	Cardiovascular system	No risk of bias information	Moderate median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	No

Xiong 2014	Active	Other objective outcome	Antineoplastic and immuno-modulating agents	Low median risk of bias	Low median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	First NRS and first RCT published in the same year	No	Yes
Xu 2016	Both active and placebo-controlled studies	Other objective outcome	Anti-infective for systemic use	No risk of bias information	No risk of bias information	2000-2009	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	No
Yang 2016	Active	Other objective outcome	Anti-infective for systemic use	Low median risk of bias	Moderate median risk of bias	2010 and later	High (score of 10-12 of 12)	NRS published before first RCT	No	No
Yang 2016	Active	Subjective outcome	Blood and blood forming organs	High median risk of bias	High median risk of bias	2000-2009	High (score of 10-12 of 12)	First RCT published before NRS	No	Yes
Yang 2016	Placebo or no treatment	Other objective outcome	Anti-infective for systemic use	Moderate median risk of bias	High median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	Yes
Yang 2016	Active	Subjective outcome	Anti-infective for systemic use	Low median risk of bias	High median risk of bias	2000-2009	High (score of 10-12 of 12)	NRS published before first RCT	No	No

Yang 2017	Active	Other objective outcome	Antineoplastic and immuno-modulating agents	High median risk of bias	Low median risk of bias	2010 and later	High (score of 10-12 of 12)	NRS published before first RCT	No	No
Yao 2016	Placebo or no treatment	Subjective outcome	Nervous system	No risk of bias information	No risk of bias information	2010 and later	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	No
Ye 2015	Placebo or no treatment	Subjective outcome	Cardiovascular system	No risk of bias information	No risk of bias information	2010 and later	Low (score of 4-6 of 12)	First RCT published before NRS	No	No
Yin 2016	Active	Other objective outcome	Antineoplastic and immuno-modulating agents	Low median risk of bias	Low median risk of bias	2010 and later	High (score of 10-12 of 12)	NRS published before first RCT	No	No
Yong 2015	Active	Other objective outcome	Antineoplastic and immuno-modulating agents	Low median risk of bias	Low median risk of bias	2010 and later	High (score of 10-12 of 12)	NRS published before first RCT	No	Yes
Yong 2017	Active	Subjective outcome	Blood and blood forming organs	No risk of bias information	No risk of bias information	2010 and later	High (score of 10-12 of 12)	NRS published before first RCT	No	Yes
Yuan 2014	Active	Subjective outcome	Nervous system	No risk of bias information	No risk of bias information	2000-2009	Moderate (score of 7-9 of 12)	NRS published	No	No

								before first RCT		
Zaiem 2017	Placebo or no treatment	Other objective outcome	Genito-urinary system and sex hormones	High median risk of bias	High median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	No
Zeng 2016	Active	Other objective outcome	Alimentary tract and metabolism	No risk of bias information	Low median risk of bias	2000-2009	High (score of 10-12 of 12)	NRS published before first RCT	No	No
Zeng 2017	Placebo or no treatment	Mortality	Blood and blood forming organs	High median risk of bias	Low median risk of bias	Before 2000	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	No
Zhai 2016	Placebo or no treatment	Subjective outcome	Nervous system	Low median risk of bias	Low median risk of bias	2000-2009	High (score of 10-12 of 12)	First NRS and first RCT published in the same year	No	No
Zhang 2014	Active	Mortality	Antineoplastic and immuno-modulating agents	High median risk of bias	High median risk of bias	Before 2000	High (score of 10-12 of 12)	First RCT published before NRS	No	No
Zhang 2014	Placebo or no treatment	Other objective outcome	Antineoplastic and immuno-	High median risk of bias	High median risk of bias	2000-2009	High (score of 10-12 of 12)	First RCT published before NRS	Yes	Yes

			modulating agents							
Zhang 2014	Placebo or no treatment	Subjective outcome	Nervous system	High median risk of bias	High median risk of bias	2010 and later	Low (score of 4-6 of 12)	NRS published before first RCT	No	No
Zhang 2016	Placebo or no treatment	Other objective outcome	Genito-urinary system and sex hormones	Moderate median risk of bias	No risk of bias information	2010 and later	High (score of 10-12 of 12)	NRS published before first RCT	No	No
Zhang 2017	Placebo or no treatment	Other objective outcome	Antineoplastic and immuno-modulating agents	Moderate median risk of bias	No risk of bias information	2000-2009	High (score of 10-12 of 12)	First NRS and first RCT published in the same year	No	No
Zhang 2017	Active	Other objective outcome	Respiratory system	High median risk of bias	Low median risk of bias	2010 and later	High (score of 10-12 of 12)	First RCT published before NRS	Yes	Yes
Zhang 2017	Active	Other objective outcome	Blood and blood forming organs	Moderate median risk of bias	Low median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	No
Zhang 2017	Active	Other objective outcome	Blood and blood forming organs	Low median risk of bias	Low median risk of bias	2010 and later	High (score of 10-12 of 12)	First RCT published before NRS	No	No

Zhang 2017	Both active and placebo-controlled studies	Subjective outcome	Systemic hormonal preparations	No risk of bias information	No risk of bias information	2000-2009	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	No
Zhang 2017	Both active and placebo-controlled studies	Other objective outcome	Alimentary tract and metabolism	Low median risk of bias	Low median risk of bias	2010 and later	Low (score of 4-6 of 12)	First RCT published before NRS	No	Yes
Zhao 2015	Placebo or no treatment	Other objective outcome	Musculo-skeletal system	Low median risk of bias	Low median risk of bias	2000-2009	High (score of 10-12 of 12)	NRS published before first RCT	No	No
Zhao 2015	Placebo or no treatment	Other objective outcome	Genito-urinary system and sex hormones	High median risk of bias	High median risk of bias	2000-2009	High (score of 10-12 of 12)	NRS published before first RCT	No	No
Zhao 2016	Placebo or no treatment	Subjective outcome	Alimentary tract and metabolism	High median risk of bias	Low median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	Yes
Zhao 2016	Placebo or no treatment	Other objective outcome	Cardiovascular system	No risk of bias information	No risk of bias information	2000-2009	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	No
Zhao 2017	Placebo or no treatment	Mortality	Antineoplastic and immuno-	Low median risk of bias	High median risk of bias	2000-2009	Low (score of 4-6 of 12)	First RCT published before NRS	No	No

			modulating agents							
Zhao 2017	Placebo or no treatment	Other objective outcome	Blood and blood forming organs	Low median risk of bias	Low median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	First NRS and first RCT published in the same year	No	No
Zhao 2018	Placebo or no treatment	Other objective outcome	Antineoplastic and immuno-modulating agents	Moderate median risk of bias	Low median risk of bias	2010 and later	High (score of 10-12 of 12)	NRS published before first RCT	No	No
Zheng 2014	Active	Subjective outcome	Dermatologicals	High median risk of bias	Moderate median risk of bias	2000-2009	High (score of 10-12 of 12)	NRS published before first RCT	Yes	Yes
Zheng 2016	Placebo or no treatment	Other objective outcome	Antineoplastic and immuno-modulating agents	High median risk of bias	No risk of bias information	2010 and later	Moderate (score of 7-9 of 12)	First NRS and first RCT published in the same year	No	No
Zheng 2017	Placebo or no treatment	Other objective outcome	Cardiovascular system	Moderate median risk of bias	Low median risk of bias	2010 and later	Moderate (score of 7-9 of 12)	NRS published before first RCT	No	No

Zhou 2014	Active	Mortality	Antineoplastic and immuno-modulating agents	Low median risk of bias	Low median risk of bias	2010 and later	High (score of 10-12 of 12)	First RCT published before NRS	No	No
Zhu 2016	Placebo or no treatment	Other objective outcome	Cardiovascular system	Moderate median risk of bias	Low median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	No
Zhu 2017	Active	Other objective outcome	Antineoplastic and immuno-modulating agents	High median risk of bias	High median risk of bias	2010 and later	High (score of 10-12 of 12)	NRS published before first RCT	No	No
Zhuang 2016	Placebo or no treatment	Subjective outcome	Cardiovascular system	Low median risk of bias	Low median risk of bias	2010 and later	Low (score of 4-6 of 12)	NRS published before first RCT	No	No
Ziff 2015	Placebo or no treatment	Other objective outcome	Cardiovascular system	Moderate median risk of bias	Moderate median risk of bias	2000-2009	Moderate (score of 7-9 of 12)	First RCT published before NRS	No	Yes
Zuo 2015	Active	Other objective outcome	Anti-infective for systemic use	Moderate median risk of bias	Low median risk of bias	2010 and later	High (score of 10-12 of 12)	NRS published before first RCT	No	No

Appendix 5: Cancer medicine indications included in chapter 5

INN	Condition	Condition categorisation	Indication	EMA marketing authorisation application number
adagrasib	Non-small cell lung cancer	Solid tumour oncology	KRAZATI as monotherapy is indicated for the treatment of adult patients with advanced non-small cell lung cancer (NSCLC) with KRAS G12C mutation and disease progression after at least one prior systemic therapy.	EMEA/H/C/006013
alectinib	Non-small cell lung cancer	Solid tumour oncology	Alecensa as monotherapy is indicated for the treatment of adult patients with ALK-positive advanced NSCLC previously treated with crizotinib.	EMEA/H/C/004164
amivantamab	Non-small cell lung cancer	Solid tumour oncology	Rybrevant as monotherapy is indicated for treatment of adult patients with advanced non small cell lung cancer (NSCLC) with activating epidermal growth factor receptor (EGFR) Exon 20 insertion mutations, after failure of platinum based therapy.	EMEA/H/C/005454
asparaginase erwinia chrysanthemi (recombinant)	Acute lymphoblastic leukemia and lymphoblastic lymphoma	Haematological oncology	Enrylaze is indicated as a component of a multi-agent chemotherapeutic regimen for the treatment of acute lymphoblastic leukaemia (ALL) and lymphoblastic lymphoma (LBL) in adult and paediatric patients (1 month and older) who developed hypersensitivity or silent inactivation to E. coli-derived asparaginase.	EMEA/H/C/005917
atezolizumab	Urothelial carcinoma	Solid tumour oncology	Tecentriq as monotherapy is indicated for the treatment of adult patients with locally advanced or	EMEA/H/C/004143

			metastatic urothelial carcinoma (UC) after prior platinum-containing chemotherapy.	
atezolizumab	Urothelial carcinoma	Solid tumour oncology	Tecentriq as monotherapy is indicated for the treatment of adult patients with locally advanced or metastatic urothelial carcinoma (UC) who are considered cisplatin ineligible.	EMEA/H/C/004143
avapritinib	Gastrointestinal Stromal Tumors	Solid tumour oncology	Ayvakyt is indicated as monotherapy for the treatment of adult patients with unresectable or metastatic gastrointestinal stromal tumours (GIST) harbouring the platelet-derived growth factor receptor alpha (PDGFRA) D842V mutation.	EMEA/H/C/005208
avelumab	Merkel cell carcinoma	Haematological oncology	Bavencio is indicated as monotherapy for the treatment of adult patients with metastatic Merkel cell carcinoma (MCC).	EMEA/H/C/004338
axicabtagene ciloleucel	Follicular Lymphoma; Diffuse Large B-Cell Lymphoma	Haematological oncology	Yescarta is indicated for the treatment of adult patients with relapsed or refractory diffuse large B cell lymphoma (DLBCL) and primary mediastinal large B cell lymphoma (PMBCL), after two or more lines of systemic therapy.	EMEA/H/C/004480
belantamab mafodotin	Multiple Myeloma	Haematological oncology	BLENREP is indicated as monotherapy for the treatment of multiple myeloma in adult patients, who have received at least four prior therapies and whose disease is refractory to at least one proteasome inhibitor, one immunomodulatory agent, and an anti-CD38 monoclonal antibody, and who have demonstrated disease progression on the last therapy.	EMEA/H/C/004935

blinatumomab	Precursor Cell Lymphoblastic Leukemia-Lymphoma	Haematological oncology	Blincyto is indicated for the treatment of adults with Philadelphia chromosome negative relapsed or refractory B-precursor acute lymphoblastic leukaemia (ALL).	EMEA/H/C/003731
brexucabtagene autoleucel	Mantle cell lymphoma (MCL)	Haematological oncology	Tecartus is indicated for the treatment of adult patients with relapsed or refractory mantle cell lymphoma (MCL) after two or more lines of systemic therapy including a Bruton's tyrosine kinase (BTK) inhibitor.	EMEA/H/C/005102
brigatinib	Non-small cell lung cancer	Solid tumour oncology	Alunbrig is indicated as monotherapy for the treatment of adult patients with anaplastic lymphoma kinase (ALK) positive advanced non-small cell lung cancer (NSCLC) previously treated with crizotinib.	EMEA/H/C/004248
capmatinib	Non-small cell lung cancer	Solid tumour oncology	Tabrecta as monotherapy is indicated for the treatment of adult patients with advanced non-small cell lung cancer (NSCLC) harbouring alterations leading to mesenchymal epithelial transition factor gene exon 14 (METex14) skipping, who require systemic therapy following prior treatment with immunotherapy and/or platinum-based chemotherapy.	EMEA/H/C/004845
cemiplimab	Cutaneous squamous cell carcinoma (CSCC)	Haematological oncology	LIBTAYO as monotherapy is indicated for the treatment of adult patients with metastatic or locally advanced cutaneous squamous cell carcinoma who are not candidates for curative surgery or curative radiation.	EMEA/H/C/004844

ceritinib	Non-small cell lung cancer	Solid tumour oncology	Zykadia is indicated for the treatment of adult patients with anaplastic lymphoma kinase (ALK) positive advanced non small cell lung cancer (NSCLC) previously treated with crizotinib.	EMEA/H/C/003819
ciltacabtagene autoleucel	Multiple Myeloma	Haematological oncology	CARVYKTI is indicated for the treatment of adult patients with relapsed and refractory multiple myeloma, who have received at least three prior therapies, including an immunomodulatory agent, a proteasome inhibitor and an anti-CD38 antibody and have demonstrated disease progression on the last therapy.	EMEA/H/C/005095
daratumumab	Multiple Myeloma	Haematological oncology	Darzalex as monotherapy is indicated for the treatment of adult patients with relapsed and refractory multiple myeloma, whose prior therapy included a proteasome inhibitor and an immunomodulatory agent and who have demonstrated disease progression on the last therapy.	EMEA/H/C/004077
dostarlimab	Endometrial cancer	Solid tumour oncology	Jemperli is indicated as monotherapy for the treatment of adult patients with mismatch repair deficient (dMMR)/microsatellite instability high (MSI H) recurrent or advanced endometrial cancer (EC) that has progressed on or following prior treatment with a platinum containing regimen.	EMEA/H/C/005204

elranatamab	Multiple Myeloma	Haematological oncology	ELREXFIO is indicated as monotherapy for the treatment of adult patients with relapsed and refractory multiple myeloma, who have received at least three prior therapies, including an immunomodulatory agent, a proteasome inhibitor, and an anti-CD38 antibody and have demonstrated disease progression on the last therapy.	EMEA/H/C/005908
entrectinib	NTRK-positive solid tumors	Solid tumour oncology	Rozlytrek as monotherapy is indicated for the treatment of adult and paediatric patients 12 years of age and older, with solid tumours that have a neurotrophic tyrosine receptor kinase (NTRK) gene fusion, - who have a disease that is locally advanced, metastatic or where surgical resection is likely to result in severe morbidity, and - who have not received a prior NTRK inhibitor - who have no satisfactory treatment options.	EMEA/H/C/004936
entrectinib	Non-small cell lung cancer	Solid tumour oncology	Rozlytrek as monotherapy is indicated for the treatment of adult patients with ROS1-positive, advanced non-small cell lung cancer (NSCLC) not previously treated with ROS1 inhibitors.	EMEA/H/C/004936
epcoritamab	B-cell lymphoma	Haematological oncology	Tepkinly as monotherapy is indicated for the treatment of adult patients with relapsed or refractory diffuse large B-cell lymphoma (DLBCL) after two or more lines of systemic therapy.	EMEA/H/C/005985

futibatinib	Cholangiocarcinoma	Solid tumour oncology	Lytgobi monotherapy is indicated for the treatment of adult patients with locally advanced or metastatic cholangiocarcinoma with a fibroblast growth factor receptor 2 (FGFR2) fusion or rearrangement that have progressed after at least one prior line of systemic therapy.	EMEA/H/C/005627
glofitamab	B-cell lymphoma	Haematological oncology	Columvi as monotherapy is indicated for the treatment of adult patients with relapsed or refractory diffuse large B-cell lymphoma (DLBCL), after two or more lines of systemic therapy.	EMEA/H/C/005751
ibrutinib	Mantle cell lymphoma (MCL)	Haematological oncology	IMBRUVICA is indicated for the treatment of adult patients with relapsed or refractory mantle cell lymphoma (MCL).	EMEA/H/C/003791
idecabtagene vicleucel	Multiple Myeloma	Haematological oncology	Abecma is indicated for the treatment of adult patients with relapsed and refractory multiple myeloma who have received at least three prior therapies, including an immunomodulatory agent, a proteasome inhibitor and an anti CD38 antibody and have demonstrated disease progression on the last therapy.	EMEA/H/C/004662
idelalisib	Non-Hodgkin Lymphoma	Haematological oncology	Zydelig is indicated as monotherapy for the treatment of adult patients with follicular lymphoma (FL) that is refractory to two prior lines of treatment.	EMEA/H/C/003843

larotrectinib	Solid tumours	Solid tumour oncology	VITRAKVI as monotherapy is indicated for the treatment of adult and paediatric patients with solid tumours that display a Neurotrophic Tyrosine Receptor Kinase (NTRK) gene fusion, - who have a disease that is locally advanced, metastatic or where surgical resection is likely to result in severe morbidity, and - who have no satisfactory treatment options.	EMEA/H/C/004919
lisocabtagene maraleucel	B-cell lymphoma; follicular lymphoma; mediastinal neoplasms	Haematological oncology	Breyanzi is indicated for the treatment of adult patients with relapsed or refractory diffuse large B-cell lymphoma (DLBCL), primary mediastinal large B-cell lymphoma (PMBCL) and follicular lymphoma grade 3B (FL3B), after two or more lines of systemic therapy.	EMEA/H/C/004731
loncastuximab tesirine	B-cell lymphoma	Haematological oncology	Zynlonta as monotherapy is indicated for the treatment of adult patients with relapsed or refractory diffuse large B-cell lymphoma (DLBCL) and high-grade B-cell lymphoma (HGBL), after two or more lines of systemic therapy.	EMEA/H/C/005685
lorlatinib	Non-small cell lung cancer	Solid tumour oncology	Lorviqua as monotherapy is indicated for the treatment of adult patients with anaplastic lymphoma kinase (ALK)-positive advanced non-small cell lung cancer (NSCLC) whose disease has progressed after: - alectinib or ceritinib as the first ALK tyrosine kinase inhibitor (TKI) therapy; or - crizotinib and at least one other ALK TKI.	EMEA/H/C/004646

melphalan flufenamide	Multiple Myeloma	Haematological oncology	Pepaxti is indicated, in combination with dexamethasone, for the treatment of adult patients with multiple myeloma who have received at least three prior lines of therapies, whose disease is refractory to at least one proteasome inhibitor, one immunomodulatory agent, and one anti-CD38 monoclonal antibody, and who have demonstrated disease progression on or after the last therapy. For patients with a prior autologous stem cell transplantation, the time to progression should be at least 3 years from transplantation.	EMEA/H/C/005681
mosunetuzumab	Follicular lymphoma	Haematological oncology	Lunsumio as monotherapy is indicated for the treatment of adult patients with relapsed or refractory follicular lymphoma (FL) who have received at least two prior systemic therapies.	EMEA/H/C/005680
moxetumomab pasudotox	Hairy cell leukaemia	Haematological oncology	PRODUCT WITHDRAWN BY MANUFACTURER POST-APPROVAL Lumoxiti as monotherapy is indicated for the treatment of adult patients with relapsed or refractory hairy cell leukaemia (HCL) after receiving at least two prior systemic therapies, including treatment with a purine nucleoside analogue (PNA).	EMEA/H/C/005322
osimertinib	Non-small cell lung cancer	Solid tumour oncology	Tagrisso is indicated for the treatment of adult patients with locally advanced or metastatic epidermal growth factor receptor (EGFR) T790M mutation-positive non-small-cell lung cancer (NSCLC).	EMEA/H/C/004124

pemigatinib	Cholangiocarcinoma	Solid tumour oncology	Pemazyre monotherapy is indicated for the treatment of adults with locally advanced or metastatic cholangiocarcinoma with a fibroblast growth factor receptor 2 (FGFR2) fusion or rearrangement that have progressed after at least one prior line of systemic therapy.	EMEA/H/C/005266
pirtobrutinib	Mantle cell lymphoma (MCL)	Haematological oncology	Jaypirca as monotherapy is indicated for the treatment of adult patients with relapsed or refractory mantle cell lymphoma (MCL) who have been previously treated with a Bruton's tyrosine kinase (BTK) inhibitor.	EMEA/H/C/005863
pralsetinib	Non-small cell lung cancer	Solid tumour oncology	Gavreto is indicated as monotherapy for the treatment of adult patients with rearranged during transfection (RET) fusion-positive advanced non-small cell lung cancer (NSCLC) not previously treated with a RET inhibitor.	EMEA/H/C/005413
rucaparib	Ovarian cancer	Solid tumour oncology	Rubraca is indicated as monotherapy treatment of adult patients with platinum sensitive, relapsed or progressive, BRCA mutated (germline and/or somatic), high-grade epithelial ovarian, fallopian tube, or primary peritoneal cancer, who have been treated with two or more prior lines of platinum based chemotherapy, and who are unable to tolerate further platinum based chemotherapy.	EMEA/H/C/004272

selinexor	Multiple Myeloma	Haematological oncology	NEXPOVIO is indicated in combination with dexamethasone for the treatment of multiple myeloma in adult patients who have received at least four prior therapies and whose disease is refractory to at least two proteasome inhibitors, two immunomodulatory agents and an anti-CD38 monoclonal antibody, and who have demonstrated disease progression on the last therapy.	EMEA/H/C/005127
selpercatinib	Non-small cell lung cancer	Solid tumour oncology	Retsevmo as monotherapy is indicated for the treatment of adults with: – advanced RET fusion positive non small cell lung cancer (NSCLC) who require systemic therapy following prior treatment with immunotherapy and/or platinum-based chemotherapy	EMEA/H/C/005375
selpercatinib	Thyroid cancer	Solid tumour oncology	Retsevmo as monotherapy is indicated for the treatment of adults with: – advanced RET fusion positive thyroid cancer who require systemic therapy following prior treatment with sorafenib and/or lenvatinib.	EMEA/H/C/005375
selpercatinib	Thyroid cancer	Solid tumour oncology	Retsevmo as monotherapy is indicated for the treatment of adults and adolescents 12 years and older with advanced RET mutant medullary thyroid cancer (MTC) who require systemic therapy following prior treatment with cabozantinib and/or vandetanib.	EMEA/H/C/005375
sonidegib	Basal Cell Carcinoma	Solid tumour oncology	Odomzo is indicated for the treatment of adult patients with locally advanced basal cell carcinoma	EMEA/H/C/002839

			(BCC) who are not amenable to curative surgery or radiation therapy.	
sotorasib	Non-small cell lung cancer	Solid tumour oncology	Lumykras as monotherapy is indicated for the treatment of adults with advanced non-small cell lung cancer (NSCLC) with KRAS G12C mutation and who have progressed after at least one prior line of systemic therapy.	EMEA/H/C/005522
tafasitamab	B-cell lymphoma	Haematological oncology	Minjuvi in combination with lenalidomide followed by Minjuvi monotherapy for the treatment of adult patients with relapsed or refractory diffuse large B-cell lymphoma (DLBCL) who are not eligible for autologous stem cell transplant (ASCT).	EMEA/H/C/005436
tagraxofusp	Blastic plasmacytoid dendritic cell neoplasm (BPDCN)	Haematological oncology	Elzonris is indicated as monotherapy for the first-line treatment of adult patients with blastic plasmacytoid dendritic cell neoplasm (BPDCN).	EMEA/H/C/005031
talquetamab	Multiple Myeloma	Haematological oncology	TALVEY is indicated as monotherapy for the treatment of adult patients with relapsed and refractory multiple myeloma, who have received at least 3 prior therapies, including an immunomodulatory agent, a proteasome inhibitor, and an anti-CD38 antibody and have demonstrated disease progression on the last therapy.	EMEA/H/C/005864

teclistamab	Multiple Myeloma	Haematological oncology	Tecvayli is indicated as monotherapy for the treatment of adult patients with relapsed and refractory multiple myeloma, who have received at least three prior therapies, including an immunomodulatory agent, a proteasome inhibitor, and an anti-CD38 antibody and have demonstrated disease progression on the last therapy.	EMEA/H/C/005865
tepotinib	Non-small cell lung cancer	Solid tumour oncology	Tepmetko as monotherapy is indicated for the treatment of adult patients with advanced non-small cell lung cancer (NSCLC) harbouring alterations leading to mesenchymal-epithelial transition factor gene exon 14 (METex14) skipping, who require systemic therapy following prior treatment with immunotherapy and/or platinum-based chemotherapy.	EMEA/H/C/005524
tisagenlecleucel	Acute lymphoblastic leukemia (ALL)	Haematological oncology	Kymriah is indicated for the treatment of: - Paediatric and young adult patients up to 25 years of age with B-cell acute lymphoblastic leukaemia (ALL) that is refractory, in relapse post-transplant or in second or later relapse.	EMEA/H/C/004090
tisagenlecleucel	Diffuse large B-cell lymphoma	Haematological oncology	Kymriah is indicated for the treatment of: - Adult patients with relapsed or refractory diffuse large B-cell lymphoma (DLBCL) after two or more lines of systemic therapy.	EMEA/H/C/004090
trastuzumab deruxtecan	Breast cancer	Solid tumour oncology	Enhertu as monotherapy is indicated for the treatment of adult patients with unresectable or	EMEA/H/C/005124

			metastatic HER2 positive breast cancer who have received two or more prior anti HER2 based regimens.	
venetoclax	Chronic Lymphocytic Leukemia (CLL)	Haematological oncology	Venclyxto monotherapy is indicated for the treatment of chronic lymphocytic leukaemia (CLL) in the presence of 17p deletion or TP53 mutation in adult patients who are unsuitable for or have failed a B-cell receptor pathway inhibitor. Venclyxto monotherapy is indicated for the treatment of CLL in the absence of 17p deletion or TP53 mutation in adult patients who have failed both chemoimmunotherapy and a B-cell receptor pathway inhibitor.	EMEA/H/C/004106

Appendix 6: Comparators for hypothetical RCTs of cancer medicines approved based on single-arm trials

Generic name	Indication	Pivotal trial name*	Primary endpoint	Comparator for a hypothetical RCT	Source for comparator	Sample size hypothetical RCT	Calculated sample fits in pivotal trial
ibrutinib	IMBRUVICA is indicated for the treatment of adult patients with relapsed or refractory mantle cell lymphoma (MCL).	PCYC-1104-CA	Overall Response Rate (ORR)	Temsirolimus	RAY trial (NCT01646021), as used in EPAR and NICE guidance. The EPAR clearly states that "In the EU, temsirolimus is the only approved treatment for relapsed or refractory MCL." (EPAR, pg 9). Note that temsirolimus is licensed in England but does not represent clinical practice in England because of its low efficacy. Nevertheless, NICE used the RAY trial since no data for other relevant treatments in England were available: "The committee understood that the clinical evidence for ibrutinib came from 1 randomised	170	No

					controlled trial (RAY), in which ibrutinib was compared with temsirolimus, and 2 single-arm studies (SPARK and PCYC-1104). The committee concluded that the studies were of a reasonable quality but were limited by the lack of a comparison against a treatment used in UK clinical practice." (NICE guidance, pg. 17) This was the confirmatory phase III trial for this conditional marketing authorisation.		
ceritinib	Zykadia is indicated for the treatment of adult patients with anaplastic lymphoma kinase (ALK) positive advanced non small cell lung cancer (NSCLC) previously treated with crizotinib.	CLDK378X2101	Overall Response Rate (ORR)	Chemotherapy (pemetrexed)	Chemotherapy (Pemetrexed) is often cited as comparator (EMA, ESMO, Djulbegovic et al, 2018, Rittberg et al, 2021). The EPAR refers to response rates observed for chemotherapy (pemetrexed) when discussing the benefits and risks of ceritinib (EPAR, pg 102). ASCEND-5 trial (NCT01828112) is mentioned in EMA and ESMO documents. This was the	28	Yes

					confirmatory phase III trial for this conditional marketing authorisation.		
idelalisib	Zydelig is indicated as monotherapy for the treatment of adult patients with follicular lymphoma (FL) that is refractory to two prior lines of treatment.	IDELA (CAL-101, 101-09)	Overall Response Rate (ORR)	NA	While there is not definitive standard of care for this indication, bendamustine was a medication most commonly referred to by regulatory documents and literature. EMA includes all prior therapies in its indication; however, most studies have used patients who have been previously treated with rituximab, which excludes that as an option.	NA	NA
blinatumomab	Blincyto is indicated for the treatment of adults with Philadelphia chromosome negative relapsed or refractory B-precursor acute	MT103-211	Overall Response Rate (ORR)	Standard of care chemotherapy	Chemotherapy widely considered best alternative therapy, including in EMA regulatory review and NICE guidance. TOWER study (NCT02013167) was also mentioned in these documents. Response rate for TOWER was defined as participants with complete	204	No

	lymphoblastic leukaemia (ALL).				remission, complete remission with hematological recovery, or complete remission with incomplete hematological recovery.The pivotal trial for blinatumomab only included complete remission and complete remission with hematological recovery.		
sonidegib	Odomzo is indicated for the treatment of adult patients with locally advanced basal cell carcinoma (BCC) who are not amenable to curative surgery or radiation therapy.	BOLT (CLDE225A2201)	Objective Response Rate (ORR)	Vismodegib	Although palliative care is most commonly used because available treatment options (surgery, chemotherapy, radiation therapy) are not effective, EMA used an indirect comparison with vismodegib in its appraisal. In the EPAR for sonidegib, vismodegib was identified as recently approved option for the treatment of "patients with symptomatic metastatic basal cell carcinoma or locally advanced basal cell carcinoma inappropriate for surgery or radiotherapy". Data on effect estimates for vismodegib were not	1560404	No

					provided in the EPAR and were instead extracted from clinicaltrials.gov for the pivotal vismodegib trial (NCT00833417).		
osimertinib	Tagrisso is indicated for the treatment of adult patients with locally advanced or metastatic epidermal growth factor receptor (EGFR) T790M mutation-positive non-small-cell lung cancer (NSCLC).	AURA (D5160C00001)	Objective Response Rate (ORR)	Platinum-based doublet chemotherapy	Platinum-based doublet chemotherapy was mentioned by NICE and ESMO, as well as by EMA under additional considerations. Data was extracted from the AURA3 trial (NCT02151981), which was the confirmatory phase III trial for this conditional marketing authorisation by EMA. The trial was underway at the time of marketing authorisation. EPAR included two pivotal trials, of which the trial with the more favourable results was used to calculate an effect sizes and compare sample size to a hypothetical RCT.	42	Yes

daratumumab	Darzalex as monotherapy is indicated for the treatment of adult patients with relapsed and refractory multiple myeloma, whose prior therapy included a proteasome inhibitor and an immunomodulatory agent and who have demonstrated disease progression on the last therapy.	GEN501	Overall Response Rate (ORR)	Pomalidomide + dexamethasone	Pomalidomide plus dexamethasone used because it was discussed as an option by both EMA and NICE. In the EPAR, two options for third-line treatment are discussed as approved agents in the EU (pomalidomide in combination with dexamethasone and panobinostat in combination with bortezomib and dexamethasone) (EPAR, pg 8). The NICE guidance states that "[a]fter 3 previous lines of treatment, pomalidomide plus dexamethasone is the only relevant comparator" (NICE guidance, pg 9). Data extracted from NIMBUS trial (NCT01311687). EPAR included two pivotal trials, of which the trial with the more favourable results was used to calculate an effect sizes and compare sample size to a hypothetical RCT.	432	No

alectinib	Alecensa as monotherapy is indicated for the treatment of adult patients with ALK-positive advanced NSCLC previously treated with crizotinib.	NP28761	Objective Response Rate (ORR)	Chemotherapy (pemetrexed + docetaxel)	Chemotherapy (pemetrexed + docetaxel) was mentioned by both EMA and ESMO. NICE was not able to publish guidance on aletinib because the company did not submit evidence for seeking reimbursement. The ALUR trial (NCT02604342) was mentioned by both EMA (in considerations section of EPAR) and ESMO. The EMA referred to the results of that trial (which was the confirmatory trial for crizotinib, another agent with conditional marketing authorisation for the same indication) when contextualising the results of the pivotal trial for alectinib and therefore clearly considered the comparator relevant. EPAR included two pivotal trials, of which the trial with the more favourable results was used to calculate an effect sizes and compare	14	Yes

					sample size to a hypothetical RCT.		
venetoclax	Venclyxto monotherapy is indicated for the treatment of chronic lymphocytic leukaemia (CLL) in the presence of 17p deletion or TP53 mutation in adult patients who are unsuitable for or have failed a B-cell receptor pathway inhibitor. Venclyxto monotherapy is indicated for the treatment of CLL	M13-982	Overall Response Rate (ORR)	Ibrutinib	Ibrutinib was commonly referenced as treatment option by EMA and in published literature. Data were extracted from the RESONATE trial (NCT01578707). Note that the indication approved by EMA was for patients who are unsuitable for or have failed a B-cell receptor pathway inhibitor, i.e. including ibrutinib. However, the pivotal trial did not specify this inclusion criteria and included all previously treated patients.	48	Yes

	in the absence of 17p deletion or TP53 mutation in adult patients who have failed both chemoimmunotherapy and a B-cell receptor pathway inhibitor.						
atezolizumab	Tecentriq as monotherapy is indicated for the treatment of adult patients with locally advanced or metastatic urothelial carcinoma (UC) after prior platinum-containing chemotherapy.	IMvigor 210 (GO29293; Cohort 2)	Objective Response Rate (ORR)	Chemotherapy (vinflunine, paclitaxel, or docetaxel)	Alternative treatment and trial (IMvigor211; NCT02302807) were mentioned by multiple regulatory and other documents (EMA, NICE, ESMO, Rittberg el al 2021, Ribeiro et al 2022). Note that the pivotal trial for atezolizumab used confirmed responses, while IMvigor211 also included unconfirmed responses.	38540	No
atezolizumab	Tecentriq as monotherapy is indicated for the treatment of adult patients with locally advanced or metastatic	IMvigor 210 (GO29293; Cohort 1)	Objective Response Rate (ORR)	Chemotherapy (gemcitabine + carboplatin/cisplatin)	Carboplatin-based regimen was mentioned in the EPAR as one of two recommended options in European treatment guidelines (EPAR, pg 11). However, in its discussion of benefits and	178	No

	urothelial carcinoma (UC) who are considered cisplatin ineligible.				risks, the EMA considered historical data for gemcitabine + carboplatin/cisplatin (EPAR, pg 196) and therefore clearly considered this the most relevant comparator. The IMvigor130 trial (NCT02807636) was mentioned in regulatory documents (EMA and NICE).		
rucaparib	Rubraca is indicated as monotherapy treatment of adult patients with platinum sensitive, relapsed or progressive, BRCA mutated (germline and/or somatic), high-grade epithelial ovarian, fallopian tube, or primary peritoneal cancer, who have been treated with two or more prior lines of platinum based	CO-338-010 (Part 2A)	Overall Response Rate (ORR)	DOXIL (liposomal doxorubicin)	Liposomal doxorubicin is an option in patients unable to withstand platinum-based chemotherapy. Other options involve combination therapy with a platinum-based chemotherapy agent, which is against the authorised indication for rucaparib. However, the EPAR only mentioned liposomal doxorubicin when discussing the benefits of rucaparib and therefore EMA clearly considered this the most relevant comparator (EPAR, pg 158) Despite olaparib being referenced as a potential comparator in the	32	Yes

	chemotherapy, and who are unable to tolerate further platinum based chemotherapy.				EPAR and other published studies, it is primarily used as a maintenance therapy after the use of platinum-based chemotherapy. Data for liposomal doxorubicin was extracted from OVA-301 trial (NCT00113607). EPAR included two pivotal trials, of which the trial with the more favourable results was used to calculate an effect sizes and compare sample size to a hypothetical RCT.		
avelumab	Bavencio is indicated as monotherapy for the treatment of adult patients with metastatic Merkel cell carcinoma (MCC).	JAVELIN Merkel 200 (EMR100070-003)	Durable Response Rate (DRR)	Carboplatin + etoposide	The authorised indication was for both previously treated and untreated patients. For the second-line treatment, no clear comparator could be identified. For first-line treatment, chemotherapy (carboplatin + etoposide) is widely referred to, including in EMA and NICE documents. Both of these documents also refered to the retrospective observational Study 100070-	74	Yes

					Obs001. Data were extracted for that study from the EPAR.		
brigatinib	Alunbrig is indicated as monotherapy for the treatment of adult patients with anaplastic lymphoma kinase (ALK) positive advanced non-small cell lung cancer (NSCLC) previously treated with crizotinib.	ALTA (AP26113-13-201)	Objective Response Rate (ORR)	Ceritinib	Ceritinib was mentioned as comparator in both EMA and NICE documents (EMA also mentioned alectinib). Data was extracted from trial ASCEND-2 (NCT01685609) since this trial had relevant outcome data for comparison with the pivotal trial for brigatinib.	278	No
tisagenlecleucel	Kymriah is indicated for the treatment of: - Paediatric and young adult patients up to 25 years of age with B-cell acute lymphoblastic leukaemia (ALL) that is refractory,	ELIANA (CCTL019B2202)	Overall Remission Rate (ORR)	Blinatumomab	While there are many potential treatments for this indication, both EMA and NICE selected blinatumomab as the main comparator. Data was extracted from trial NCT01471782, a phase I/II study.	32	Yes

	in relapse post-transplant or in second or later relapse.						
tisagenlecleucel	Kymriah is indicated for the treatment of: - Adult patients with relapsed or refractory diffuse large B-cell lymphoma (DLBCL) after two or more lines of systemic therapy.	JULIET (CCTL019C2201)	Overall Response Rate (ORR)	Chemoimmunotherapy	The NICE appraisal was suspended. Only EMA provided potential comparators (through an indirect comparison to the pivotal trial for tisagenlecleucel using a historical dataset). Refering to the indirect comparison, EMA clearly indicated a relevant comparator. While three studies were discussed by EMA (CORAL, PIX-310, SCHOLAR-1), CORAL (NCT00137995) was referred to as the most relevant comparator trial. Data was extracted from the EPAR for tisagenlecleucel since no results were posted on clinicaltrials.gov.	472	No

axicabtagene ciloleucel	Yescarta is indicated for the treatment of adult patients with relapsed or refractory diffuse large B cell lymphoma (DLBCL) and primary mediastinal large B cell lymphoma (PMBCL), after two or more lines of systemic therapy.	ZUMA-1	Overall Response Rate (ORR)	Salvage chemotherapy	Salvage chemotherapy and the retrospective SCHOLAR-1 study were cited by both EMA and NICE. While no single treatment was identified as standard of care, the use of an indirect comparison indicated a clear comparator (i.e. salvage chemotherapy).	36	Yes
moxetumomab pasudotox	PRODUCT WITHDRAWN BY MANUFACTURER POST-APPROVAL Lumoxiti as monotherapy is indicated for the treatment of adult patients with relapsed or refractory hairy cell leukaemia (HCL) after receiving at least	CD-ON-CAT-8015-1053	Durable Complete Response	NA	No suitable comparator identified since treatment options discussed in the EPAR and in NICE guidance were contradictory.	NA	NA

	two prior systemic therapies, including treatment with a purine nucleoside analogue (PNA).						
cemiplimab	LIBTAYO as monotherapy is indicated for the treatment of adult patients with metastatic or locally advanced cutaneous squamous cell carcinoma who are not candidates for curative surgery or curative radiation.	R2810-ONC-1540 (Study 1540)	Objective Response Rate (ORR)	NA	No suitable comparator identified since treatment options discussed in the EPAR and in NICE guidance were contradictory.	NA	NA
lorlatinib	Lorviqua as monotherapy is indicated for the treatment of adult patients with anaplastic lymphoma kinase (ALK)-positive	B7461001	Objective Response Rate (ORR)	Singlet chemotherapy	The next best treatment in this population is platinum-based doublet chemotherapy (PDC) or ACBP (Atezolizumab in combination with bevacizumab, paclitaxel and carboplatin) according to	42	Yes

	advanced non-small cell lung cancer (NSCLC) whose disease has progressed after: - alectinib or ceritinib as the first ALK tyrosine kinase inhibitor (TKI) therapy; or - crizotinib and at least one other ALK TKI.				NICE. EMA also mentioned chemotherapy as fall-back standard of care. Clinical data for a population similar to the authorised indication are, however, only available for singlet chemotherapy. These data were referenced in the EPAR and they were also submitted to NICE. Both EMA and NICE referenced the ALUR trial (NCT02604342), therefore data was extracted for this trial.		
larotrectinib	VITRAKVI as monotherapy is indicated for the treatment of adult and paediatric patients with solid tumours that display a Neurotrophic Tyrosine Receptor Kinase (NTRK) gene fusion, - who have a disease that is locally advanced,	NAVIGATE (LOXO-TRK-15002)	Overall Response Rate (ORR)	NA	Larotrectinib is intended to be used as a last-line therapy - thus, there is no standard of care. The intention of therapy in this setting is palliative. The NICE guidance used a blended arm approach submitted by the company to reflect the multiple clinical options that larotrectinib would replace. No outcome data for the blended comparator could be extracted.	NA	NA

	metastatic or where surgical resection is likely to result in severe morbidity, and - who have no satisfactory treatment options.						
tagraxofusp	Elzonris is indicated as monotherapy for the first-line treatment of adult patients with blastic plasmacytoid dendritic cell neoplasm (BPDCN).	STML-401-0114	Objective Response Rate (ORR)	NA	Regulatory documents refer to chemotherapy as the best comparative option but without specifying a particular regimen, therefore no outcome data for a cmparator could be extracted.	NA	NA
selinexor	NEXPOVIO is indicated in combination with dexamethasone for the treatment of multiple myeloma in adult patients who have received at least four prior	STORM (KCP-330-012)	Overall Response Rate (ORR)	Best supportive care	Both EMA and NICE documents mention that there are currently no preferred treatments for this indication, so best supportive care was determined to be the best comparator. The trial used was the retrospective MAMMOTH study,	7864	No

	therapies and whose disease is refractory to at least two proteasome inhibitors, two immunomodulatory agents and an anti-CD38 monoclonal antibody, and who have demonstrated disease progression on the last therapy.				mentioned by NICE. Outcome data were extracted for the group of patients with triple-refractory and quad-refractory multiple myeloma (i.e. refractory to 1 CD38-targeted monoclonal antibody therapy (CD38 MoAB) + 1 proteasome inhibitor (PI) + 1 or 2 immunomodulatory drugs (IMiD), or to 1 CD38 MoAB + 1 or 2 PIs + 1 IMiD). (Gandhi et al. Outcomes of patients with multiple myeloma refractory to CD38-targeted monoclonal antibody therapy. Leukemia. 2019 Sep;33(9):2266-2275. doi: 10.1038/s41375-019-0435-7).		

entrectinib	Rozlytrek as monotherapy is indicated for the treatment of adult and paediatric patients 12 years of age and older, with solid tumours that have a neurotrophic tyrosine receptor kinase (NTRK) gene fusion, - who have a disease that is locally advanced, metastatic or where surgical resection is likely to result in severe morbidity, and - who have not received a prior NTRK inhibitor - who have no satisfactory treatment options.	STARTRK-2 (NTRK cohort)	Objecti ve Respons e Rate (ORR)	NA	EMA discussed larotrectinib as a potential comparator, but at the time of assessment, this only had a conditional marketing authorisation and thus its clinical effectiveness could not be fully evaluated. NICE recommended best supportive care as the most appropriate comparator. In its assessment of OS and PFS, it relied on a blended arm of best supportive care, chemotherapy, and hormone therapy created by the company. While OS and PFS data exist for this blended arm, no reponse outcomes were provided, and therefore no effect size could be calculated.	NA	NA

entrectinib	Rozlytrek as monotherapy is indicated for the treatment of adult patients with ROS1-positive, advanced non-small cell lung cancer (NSCLC) not previously treated with ROS1 inhibitors.	STARTRK-2 (ROS1 cohort)	Objective Response Rate (ORR)	Crizotinib	Crizotinib was mentioned in EMA documents (referencing ESMO guidelines) as best comparator, and was also mentioned in the FDA regulatory documents for the corresponding indication. NICE also mentioned crizotinib, but did not use this as comparator in line with its policy on products only available through the Cancer Drugs Fund (CDF) and not through routine commissioning (which was the case for crizotinib). Both EMA and FDA referenced the PROFILE-1001 trial (NCT00585195) for crizotinib data. Note that there were three pivotal trials for this indication. However, the EPAR provided the pooled ORR and the number of participants it was based on, and this was used to calculate the effect size. The largest of the three pivotal trials was used to compare	21470	No

					the sample size of a hypothetical RCT.		
trastuzumab deruxtecan	Enhertu as monotherapy is indicated for the treatment of adult patients with unresectable or metastatic HER2 positive breast cancer who have received two or more prior anti HER2 based regimens.	DESTINY-Breast01 (DS8201-A-U201)	Objective Response Rate (ORR)	NA	No suitable comparator identified since treatment options discussed in the EPAR and in NICE guidance were contradictory.	NA	NA
avapritinib	Ayvakyt is indicated as monotherapy for the treatment of adult patients with unresectable or	NAVIGATOR (BLU-285-1101)	Objective Response Rate (ORR)	Imatinib	Imatinib was referenced in both EMA and NICE documents as first-line treatment for the authorised indication. Outcome data was extracted from a	0	Yes

	metastatic gastrointestinal stromal tumours (GIST) harbouring the platelet-derived growth factor receptor alpha (PDGFRA) D842V mutation.				restrospective study referenced in EMA documents (Yoo et al. Efficacy of Imatinib in Patients with Platelet-Derived Growth Factor Receptor Alpha–Mutated Gastrointestinal Stromal Tumors. Cancer Research and Treatment : Official Journal of Korean Cancer Association 2016;48(2):546-552. DOI: https://doi.org/10.4143/crt.2015.015).		
pemigatinib	Pemazyre monotherapy is indicated for the treatment of adults with locally advanced or metastatic cholangiocarcinoma with a fibroblast growth factor receptor 2 (FGFR2) fusion or rearrangement that have progressed after at	FIGHT-202 (INCB 54828-202)	Objective Response Rate (ORR)	mFOLFOX (modified folinic acid, 5-fluorouracil and oxaliplatin) +ASC (active symptom control)	This comparator was mentioned by EMA and NICE, as well as in a previous study identifying comparators for drugs authorised based on single-arm trials (Ribeiro et al, 2022). Outcome data were extracted from the publication of a phase III RCT (ABC-06 (NCT01926236)) since no results were available on clinicaltrials.gov (Lamarca et al. Second-line FOLFOX	42	Yes

	least one prior line of systemic therapy.				chemotherapy versus active symptom control for advanced biliary tract cancer (ABC-06): a phase 3, open-label, randomised, controlled trial. Lancet Oncol. 2021 May;22(5):690-701. doi: 10.1016/S1470-2045(21)00027-9).		
capmatinib	Tabrecta as monotherapy is indicated for the treatment of adult patients with advanced non-small cell lung cancer (NSCLC) harbouring alterations leading to mesenchymal epithelial transition factor gene exon 14 (METex14) skipping, who require systemic therapy following prior treatment with immunotherapy	GEOMETRY mono-1 (CINC280A2201)	Overall Response Rate (ORR)	Docetaxel	Docetaxel was clearly identified as comparator by the EMA and was also mentioned by the FDA in its assessment of the corresponding indication. NICE did not publish guidance because the company did not submit data seeking for reimbursement. Outcome data were extracted from the CheckMate057 trial (NCT01673867) referenced in regulatory documents.	58	Yes

	and/or platinum-based chemotherapy.						
selpercatinib	Retsevmo as monotherapy is indicated for the treatment of adults with: – advanced RET fusion positive non small cell lung cancer (NSCLC) who require systemic therapy following prior treatment with immunotherapy and/or platinum-based chemotherapy	LIBRETTO-001 (LOXO-RET-17001, RET fusion positive NSCLC cohort)	Overall Response Rate (ORR)	Docetaxel	Docetaxel was identified by EMA as one of several treatment options. Docetaxel was identified as main comparator in the NICE guidance. Outcome data were extracted from the REVEL trial (NCT01168973).	22	Yes
selpercatinib	Retsevmo as monotherapy is indicated for the treatment of	LIBRETTO-001 (LOXO-RET-17001, RET fusion positive	Objective Respons	Best supportive care	Both EMA and NICE discuss that if RET-fusion positive thyroid cancer progresses after sorafenib and/or	4	Yes

	adults with: – advanced RET fusion positive thyroid cancer who require systemic therapy following prior treatment with sorafenib and/or lenvatinib.	thyroid cancer cohort)	e Rate (ORR)		lenvatinib, there are not further alternative treatments. Thus, best supportive care is the best available comparator. In the NICE guidance, data for best supportive care were approximated through results from the placebo arm of a phase III RCT (SELECT; NCT01321554).		
selpercatinib	Retsevmo as monotherapy is indicated for the treatment of adults and adolescents 12 years and older with advanced RET mutant medullary thyroid cancer (MTC) who require systemic therapy following prior treatment with cabozantinib and/or vandetanib.	LIBRETTO-001 (LOXO-RET-17001, RET mutant MTC cohort)	Overall Response Rate (ORR)	Best supportive care	Both EMA and NICE discuss that if RET mutant medullary thyroid cancer (MTC) progresses after sorafenib and/or lenvatinib, there are not further alternative treatments. Thus, best supportive care is the best available comparator. In the NICE guidance, data for best supportive care were approximated through results from the placebo arm of a phase III RCT (EXAM; NCT00704730).	8	Yes

brexucabtagene autoleucel	Tecartus is indicated for the treatment of adult patients with relapsed or refractory mantle cell lymphoma (MCL) after two or more lines of systemic therapy including a Bruton's tyrosine kinase (BTK) inhibitor.	ZUMA-2 (KTE-C19-102)	Objective Response Rate (ORR)	NA	Multiple treatment options were discussed in the EPAR, and NICE guidance did not provide a clear preferred comparator either since this was still in development at the time.	NA	NA
tafasitamab	Minjuvi in combination with lenalidomide followed by Minjuvi monotherapy for the treatment of adult patients with relapsed or refractory diffuse large B-cell lymphoma (DLBCL) who are not eligible for autologous stem	L-MIND (MOR208C203)	Objective Response Rate (ORR)	Polatuzumab vedotin in combination with bendamustine + rituximab	Polatuzumab vedotin with bendamustine and rituximab was identified as best comparator as this was referenced by EMA (although at the time of the assessment only conditionally authorised) and selected as main comparator in the NICE guidance. Outcome data were extracted from a trial (NCT02257567) which was referenced in the NICE documents.	376	No

	cell transplant (ASCT).						
belantamab mafodotin	BLENREP is indicated as monotherapy for the treatment of multiple myeloma in adult patients, who have received at least four prior therapies and whose disease is refractory to at least one proteasome inhibitor, one immunomodulatory agent, and an anti-CD38 monoclonal antibody, and who have demonstrated disease progression on the last therapy.	DREAMM-2 (205678)	Overall Response Rate (ORR)	Pomalidomide + dexamethasone	Pomalidomide plus dexamethasone was identified as comparator because it was mentioned as an option by EMA, NICE, and in literature reviews. Other options (daratumumab monotherapy and selinexor monotherapy) were also referenced. However, EMA referred to treatment effects for pomalidomide + dexamethasone when contextualising the results of the pivotal trial for belantamab mafodotin (EPAR, pg 87), indicating that this was considered the most relevant comparator. Outcome data were extracted from the NIMBUS trial (NCT01311687) and literature for this trial was also consulted (Raedler et al.	542	No

					Pomalyst (Pomalidomide) Received a New Indication for Patients with Relapsed and/or Refractory Multiple Myeloma. Am Health Drug Benefits. 2016 Mar;9(Spec Feature):111-4).		
pralsetinib	Gavreto is indicated as monotherapy for the treatment of adult patients with rearranged during transfection (RET) fusion-positive advanced non-small cell lung cancer (NSCLC) not previously treated with a RET inhibitor.	ARROW (BLU-667-1101)	Overall Response Rate (ORR)	Chemoimmunotherapy (pembrolizumab plus pemetrexed and chemotherapy)	Several possible treatment options were referenced by EMA and NICE, including some that were referenced across both: single-agent chemotherapy/docetaxel, platinum-based doublet chemotherapy, and chemoimmunotherapy. According to ESMO guidelines (as referenced in the EPAR, pg 10), first-line treatment for NSCLC (irrespective of RET mutation) would be platinum doublet-based cytotoxic chemotherapy and/or immunotherapy with checkpoint inhibitor. The latter was selected as the	266	No

					most relevant comparator, since results for this (although not exclusively in patients with RET mutation) were used to contextualise results for pralsetinib by EMA (EPAR pg 135). Outcome data were extracted from trial KEYNOTE-189 (NCT02578680), which was referenced in both EMA and NICE documents.		
tepotinib	Tepmetko as monotherapy is indicated for the treatment of adult patients with advanced non-small cell lung cancer (NSCLC) harbouring alterations leading to mesenchymal-epithelial transition factor gene exon 14 (METex14) skipping, who require systemic	VISION (MS200095-0022)	Objective Response Rate (ORR)	NA	EMA and NICE documents refer to a common mix of chemotherapy and immunotherapy options, but without specifying a clear comparator. Additionally, the indications slightly differ between regulations (ex. EMA approved tepotinib as second-line treatment while NICE reviewed tepotinib as first-line treatment).	NA	NA

	therapy following prior treatment with immunotherapy and/or platinum-based chemotherapy.						
lisocabtagene maraleucel	Breyanzi is indicated for the treatment of adult patients with relapsed or refractory diffuse large B-cell lymphoma (DLBCL), primary mediastinal large B-cell lymphoma (PMBCL) and follicular lymphoma grade 3B (FL3B), after two or more lines of systemic therapy.	TRANSCEND-NHL-001 (017001, DLBCL cohort only)	Overall Response Rate (ORR)	NA	There were some differences in the indications reviewed by EMA and NICE. Both discussed chemotherapy as a potential comparator, but did not specify which treatment was best.	NA	NA

melphalan flufenamide	Pepaxti is indicated, in combination with dexamethasone, for the treatment of adult patients with multiple myeloma who have received at least three prior lines of therapies, whose disease is refractory to at least one proteasome inhibitor, one immunomodulatory agent, and one anti-CD38 monoclonal antibody, and who have demonstrated disease progression on or after the last therapy. For patients with a prior autologous stem cell transplantation,	HORIZON (OP-106)	Overall Response Rate (ORR)	Pomalidomide + dexamethasone	NICE guidance was terminated. However, a clear comparator was identified by EMA: The EPAR refers to trial OP-103 (OCEAN study; NCT03151811) as the confirmatory phase III trial for the conditional marketing authorisation of melphalan flufenamide and identified the study's control (Pomalidomide + dexamethasone) as an appropriate comparator. EMA also mentions three potential comparators (Blenrep, Nexpovio, Abecma) but at the time of assessment, all three had conditional marketing authorisations and were thus not applicable. This RCT was included as supportive study in the EMA application and was referenced in the EMA approval in support of a CMA as indication that comprehensive data will be provided.	1406	No

	the time to progression should be at least 3 years from transplantation.						
idecabtagene vicleucel	Abecma is indicated for the treatment of adult patients with relapsed and refractory multiple myeloma who have received at least three prior therapies, including an immunomodulatory agent, a proteasome inhibitor and an anti CD38 antibody and have demonstrated	KarMMa (BB2121-MM-001)	Overall Response Rate (ORR)	Standard regimen (combinations of immunomodulatory agents, PIs, monoclonal antibodies, corticosteroids, and cytotoxic agents)	No single standard of care therapy exists, as treatments are selected based on prior failed regimens. However, EMA clearly considered standard regimens and their effectiveness when assessing ide-cel, as evidenced by (1) the description of the treatment landscape and ESMO recommendations (triplet regimen based on a backbone of pomalidomide and dexamethasone, daratumumab monotherapy or combination, or enrolment in a clinical trial, EPAR pg 14), (2) the use of a	38	Yes

	disease progression on the last therapy.				real-world study (NDS-MM-003) with treatments including combinations of immunomodulatory agents, PIs, monoclonal antibodies, corticosteroids, and cytotoxic agents to contextualise treatment effects for ide-cel obtained from the pivotal trial, and (3) the granting of a conditional marketing authorisation for ide-cel contingent on conducting a comparative phase III trial (KarMMa-003, NCT03651128) which compared ide-cel to "standard regimens as per Investigator's discretion" in the authorised indication. Outcome data were extracted from the observational NDS-MM-003 study, as this was available at the time of assessment by the EMA. NICE guidance was terminated. While a final scope with two potential comparators was available,		

					there was no explanatory text for these.		
dostarlimab	Jemperli is indicated as monotherapy for the treatment of adult patients with mismatch repair deficient (dMMR)/microsatellite instability high (MSI H) recurrent or advanced endometrial cancer (EC) that	GARNET (4010-01-001)	Objective Response Rate (ORR)	NA	EMA and NICE documents describe different comparator options. NICE used a trial with multiple comparators in control group but none of the comparators match those referred to by EMA.	NA	NA

	has progressed on or following prior treatment with a platinum containing regimen.						
loncastuximab tesirine	Zynlonta as monotherapy is indicated for the treatment of adult patients with relapsed or refractory diffuse large B-cell lymphoma (DLBCL) and high-grade B-cell lymphoma (HGBL), after two or more lines of systemic therapy.	LOTIS-2 (ADCT-402-201)	Overall Response Rate (ORR)	Polatuzumab vedotin in combination with bendamustine + rituximab	Polatuzumab vedotin (Polivy) in combination with bendamustine + rituximab was mentioned as common treatment for this indication by both EMA and NICE. Outcome data were extracted from trial GO29365 (NCT02257567), which is used in NICE's guidance and was the pivotal trial for polatuzumab vedotin.	2324	No
amivantamab	Rybrevant as monotherapy is indicated for treatment of adult patients with advanced non small cell lung cancer (NSCLC)	CHRYSALIS (61186372EDI1001)	Overall Response Rate (ORR)	Docetaxel + ramucirumab	The combination of docetaxel and ramucirumab was mentioned by EMA. There are multiple treatment options, and NICE used a blended comparator in its guidance. However, no outcome data could be	332	No

	with activating epidermal growth factor receptor (EGFR) Exon 20 insertion mutations, after failure of platinum based therapy.				extracted for the blended comparator. In addition, FDA also referred to docetaxel and ramucirumab in its assessment of the corresponding indication. Outcome data were extracted from the REVEL trial (NCT01168973), which was referred to by EMA.		
sotorasib	Lumykras as monotherapy is indicated for the treatment of adults with advanced non-small cell lung cancer (NSCLC) with KRAS G12C mutation and who have progressed after at least one prior line of systemic therapy.	CodeBreak 100 (20170543)	Objective Response Rate (ORR)	Taxane chemotherapy (docetaxel monotherapy)	Taxane chemotherapy was referenced as comparator by both EMA and NICE. EMA generally refers to taxane chemotherapy while NICE specifically refers to docetaxel. EMA granted conditional marketing authorisation for sotorasib with a confirmatory phase III RCT comparing sotorasib to docetaxel (CodeBreaK 200 / Study 20190009), therefore this was used as comparator. Outcome data were extracted from the SELECT-1 trial (NCT01933932).	180	No

asparaginase erwinia chrysanthemi (recombinant)	Enrylaze is indicated as a component of a multi-agent chemotherapeutic regimen for the treatment of acute lymphoblastic leukaemia (ALL) and lymphoblastic lymphoma (LBL) in adult and paediatric patients (1 month and older) who developed hypersensitivity or silent inactivation to E. coli-derived asparaginase.	JZP458-201	Overall Response Rate (ORR)	NA	Regulatory bodies in both Europe and the US (EMA and FDA) referred to erwinase as the sole alternative for this indication. However, erwinase has manufacturing issues, and crisantapase is expected to fill this need.	NA	NA
ciltacabtagene autoleucel	CARVYKTI is indicated for the treatment of adult patients with relapsed and refractory multiple myeloma, who have received at least three prior therapies,	CARTITUDE-1	Overall Response Rate (ORR)	Pomalidomide + dexamethasone	EMA and NICE both discussed several treatment options (note that NICE guidance was suspended, but draft guidance was available). Pomalidomide / low-dose dexamethasone was the only therapy mentioned by both EMA and NICE. Outcome data were	6	Yes

	including an immunomodulatory agent, a proteasome inhibitor and an anti-CD38 antibody and have demonstrated disease progression on the last therapy.				extracted from the NIMBUS trial (MM-003, NCT01311687), which was referenced in the EPAR.		
futibatinib	Lytgobi monotherapy is indicated for the treatment of adult patients with locally advanced or metastatic cholangiocarcinoma with a fibroblast growth factor receptor 2 (FGFR2) fusion or rearrangement that have progressed after at least one prior line of systemic therapy.	TAS-120-101	Objective Response Rate (ORR)	Pemigatinib	Pemigatinib was mentioned as comparator by both EMA and NICE. Outcome data were extracted from the FIGHT-202 trial (NCT02924376), which was referenced in the NICE guidance.	3370	No

adagrasib	KRAZATI as monotherapy is indicated for the treatment of adult patients with advanced non-small cell lung cancer (NSCLC) with KRAS G12C mutation and disease progression after at least one prior systemic therapy.	KRYSTAL-1 (Study 849-001)	Objective Response Rate (ORR)	Docetaxel	Docetaxel monotherapy, docetaxel/nintedanib, and sotorasib (at the time with a conditional marketing authorisation) were referenced by both EMA and NICE. However, EMA requested additional data from a clinical trial (KRYSTAL-12, NCT04685135) that uses docetaxel monotherapy as a control. Hence, docetaxel monotherapy was selected as most appropriate comparator. Outcome data were extracted from the most recent completed trial referenced in the EPAR, CodeBreak200 (NCT04303780). Outcome data were extracted from the EPAR since not all relevant data were available from clinicaltrials.gov.	70	Yes
mosunetuzumab	Lunsumio as monotherapy is indicated for the treatment of adult patients with	GO29781	Overall Response Rate (ORR)	Rituximab + lenalidomide	No single standard of care could be identified. EMA and NICE referenced different options, but rituximab and lenalidomide combination	8586	No

	relapsed or refractory follicular lymphoma (FL) who have received at least two prior systemic therapies.				therapy was referenced by both. Outcome data were extracted from the AUGMENT trial (NCT01938001). EMA referenced another trial (CELESTIMO), but this trial did not yet have results.		
pirtobrutinib	Jaypirca as monotherapy is indicated for the treatment of adult patients with relapsed or refractory mantle cell lymphoma (MCL) who have been previously treated with a Bruton's tyrosine kinase (BTK) inhibitor.	LOXO-BTK-18001 (BRUIN)	Overall Response Rate (ORR)	NA	Both BTK inhibitors and brexucabtagene autoleucel are considered alternative treatments for this indication, but a specific choice is not provided in reviewed documents.	NA	NA

epcoritamab	Tepkinly as monotherapy is indicated for the treatment of adult patients with relapsed or refractory diffuse large B-cell lymphoma (DLBCL) after two or more lines of systemic therapy.	EPCORE NHL-1 (GCT3013-01; aNHL cohort only)	Overall Response Rate (ORR)	R-GemOx (rituximab + gemcitabine + oxaliplatin)	EMA and NICE documents share three potential therapies as comparators (Yescarta, Polivy+BR, R-chemotherapy). In the EMA conclusion section, an ongoing clinical trial (NCT04628494: the confirmatory phase III trial for conditional marketing authorisation of epcoritamab) is referenced that uses investigator choice of treatment as a control group (R-GemOx and bendamustine/rituximab). Hence, R-GemOx was identified as most appropriate comparator. Outcome data were extracted from the group of patients refractory to second- or later line from the retrospective SCHOLAR-1 study (Crump et al. Outcomes in refractory diffuse large B-cell lymphoma: results from the international SCHOLAR-1 study. Blood 2017; 130 (16): 1800–1808. doi:	52	Yes

					https://doi.org/10.1182/blood-2017-03-769620).		
glofitamab	Columvi as monotherapy is indicated for the treatment of adult patients with relapsed or refractory diffuse large B-cell lymphoma (DLBCL), after two or more lines of systemic therapy.	NP30179	Complete Response Rate	Polatuzumab vedotin in combination with bendamustine + rituximab	EMA and NICE documents share three potential therapies as comparators (Yescarta, Polivy+BR, R-chemotherapy). In the EMA conclusion section, only two are compared to glofitamab (Yescarta and Polivy+BR). The EMA solely discusses logistical challenges when comparing glofitamab with CAR-T, whereas it discusses clinical/surrogate endpoints when comparing glofitamab with Polivy+BR. Hence, Polivy+BR was identified as the most relevant comparator. Outcome data was extracted from the	714	No

					GO29365 trial (NCT02257567).		
talquetamab	TALVEY is indicated as monotherapy for the treatment of adult patients with relapsed and refractory multiple myeloma, who have received at least 3 prior therapies, including an immunomodulatory agent, a proteasome inhibitor, and an anti-CD38 antibody and have demonstrated disease progression on the last therapy.	MonumenTAL-1 (64407564MMY1001)	Overall Response Rate (ORR)	Pomalidomide + dexamethasone	EMA and NICE discussed several potential comparators, and the EMA document provided outcome data for different treatment options. Two comparators were commonly referenced by both: pomalidomide + low-dose dexamethasone and selinexor + dexamethasone. In its assessment of the landscape of approved treatments in the heavily pre-treated and refractory population, the EMA document concluded that the response rate was approximately 30%. Since several options exist, the comparator that most closely approximated the assumed overall response rate in this population was selected to extract outcome data, which	22	Yes

					was pomalidomide + low-dose dexamethasone according to the overview of treatment options provided by EMA. Note that the appr. 30% response rate was based on ivestigator assessment. For the calculation of an effect size, IRC-based response rate was used since this was also available from the pivotal trial for talquetamab. Data were extracted from the NIMBUS trial (NCT01311687).		
teclistamab	Tecvayli is indicated as monotherapy for the treatment of adult patients with relapsed and refractory multiple myeloma, who have received at least three prior therapies, including an immunomodulatory agent, a proteasome	MajesTEC-1 (64007957MMY1001)	Overall Response Rate (ORR)	Pomalidomide + dexamethasone	Pomalidomide + low-dose dexamethasone was mentioned by both EMA and NICE. Outcome data were extracted from the NIMBUS trial (NCT01311687).	44	Yes

	inhibitor, and an anti-CD38 antibody and have demonstrated disease progression on the last therapy.						
elranatamab	ELREXFIO is indicated as monotherapy for the treatment of adult patients with relapsed and refractory multiple myeloma, who have received at least three prior therapies, including an immunomodulatory agent, a proteasome inhibitor, and an anti-CD38 antibody and have demonstrated disease progression on the last therapy.	MagnetisMM-3 (C107003)	Objective Response Rate (ORR)	NA	EPAR and NICE guidance use different comparators.	NA	NA

Legend: * Note that some indications had more than one pivotal trial. For calculating effect estimates and comparing sample sizes to a hypothetical RCT, the pivotal trial with the most beneficial effect estimate for the new medicine was selected.