BJCP British Journal of Clinical Pharmacology

Bayesian analysis of real-world data as evidence for drug approval: remembering Sir Michael Rawlins

Journal:	*British Journal of Clinical Pharmacology*
Manuscript ID	ED-00385-23.R1
Manuscript Type:	Editorial
Date Submitted by the Author:	n/a
Complete List of Authors:	Szigeti, Balázs ; Imperial College London Phillips, Lawrence; London School of Economics and Political Science, Management Nutt, David; Imperial College London
Key Words:	Health Policy < Clinical Pharmacology, Biostatistics < Statistics and Study Design, Randomised controlled trial < Drug Development, Public Health
Abstract:	The two pillars of modern medical research are null hypothesis significance testing (NHST) and randomized controlled trials (RCTs), where in most RCTs the active treatment is compared to placebo. A recent expert consensus survey endorsed the statement that "Results from placebo-controlled trials are more reliable than results from any other study design." (Similon et al., 2022), reflecting that placebo controlled RCTs are considered to be the gold standard. In his Harvein Oration, a prestigious annual lecture held at the Royal College of Physicians of London, Sir Michael Rawlins, the ex-head of National Institute for Health and Care Excellence (NICE) and the Medicines and Healthcare products Regulatory Agency (MHRA), pointed out RCTs are not the apex of evidence, but rather a piece of a larger evidence puzzle: "RCTs are often called the 'gold standard' for demonstrating (or refuting) the benefits of a particular intervention. Yet the technique has important limitations of which four are particularly troublesome: the null hypothesis, probability, generalisability and resource implications" (Rawlins, 2008). Here, we follow the footsteps of Sir Michael Rawlins and highlight how the combination of real-world evidence and Bayesian analysis could complement the traditional approach of RCTs and NHST.

Bayesian analysis of real-world data as evidence for drug approval: remembering Sir Michael Rawlins

Balázs Szigeti[1*], Lawrence D Phillips[2], David Nutt[1]

[1] Centre for Psychedelic Research, Division of Psychiatry, Imperial College London, UK

[2] London School of Economics and Political Science, UK

[*] Corresponding author: balazs.szigeti07@imperial.ac.uk

This paper is written to commemorate Sir Michael Rawlins' contribution to clinical trial design and analysis – one of the many great contributions he made to the discipline of clinical pharmacology and therapeutics in his long career.

The two pillars of modern medical research are null hypothesis significance testing (NHST) and randomized controlled trials (RCTs), where in most RCTs the active treatment is compared to placebo. A recent expert consensus survey endorsed the statement that *"Results from placebo-controlled trials are more reliable than results from any other study design."* (Similon et al., 2022), reflecting that placebo controlled RCTs are considered to be the gold standard. In his Harvein Oration, a prestigious annual lecture held at the Royal College of Physicians of London, Sir Michael Rawlins, the ex-head of National Institute for Health and Care Excellence (NICE) and the Medicines and Healthcare products Regulatory Agency (MHRA), pointed out RCTs are not the apex of evidence, but rather a piece of a larger evidence puzzle: *"RCTs are often called the 'gold standard' for demonstrating (or refuting) the benefits of a particular intervention. Yet the technique has important limitations of which four are particularly troublesome: the null hypothesis, probability, generalisability and resource implications"* (Rawlins, 2008). Here, we follow the footsteps of Sir Michael Rawlins and highlight how the combination of real-world evidence and Bayesian analysis could complement the traditional approach of RCTs and NHST.

Many medical innovations crash of the rocks "of not having RCT data" with consequent suffering to patients and their families. Here we look at two cases where classic RCT data is lacking, but where real-world data already provides considerable evidence:

- Treatment of treatment-resistant epilepsy in children with medical cannabis (Zafar et al., 2020, 2021), where we define 'treatment success' as having fewer seizures.

- Treatment of treatment-resistant depression with psilocybin (Carhart-Harris et al., 2016), where we define 'treatment success' as a 50% or larger reduction of depression scores relative to baseline.

Note, we use binary outcome data here to simplify the analysis, however the arguments presented are equally applicable to continuous outcome measures. Assuming a flat prior, i.e. prior to administering the treatment equal belief that the treatment will be success / failure, it is possible to derive the probability of success when treating the next patients as

$$p(next\ patient\ success) = \frac{number\ of\ prior\ successes + 1}{number\ of\ prior\ successes + number\ of\ prior\ failures + 2}$$

and the credible interval, the Bayesian analogue of the confidence interval, can also be calculated from the appropriate beta distribution (Kruschke, 2014).

What does Bayesian analysis reveal about the medical cannabis for childhood epilepsy? In the dataset all 20 children experienced a reduction in seizure numbers, so after updating the flat prior (all success rates between 0 and 100% are possible), the probability that the next patient will improve is 95% with, a 95% credible interval of 87-100%. Even if we choose a sceptical prior, a gentle distribution excluding 0 and 100%, with a mean at 25%, the probability that the next patient will improve is still 88% with a 95% credible interval of 75%-98%.

Similar results are obtained for treating treatment resistant depression with psilocybin. In this case results depend on which exact depression measure is used. The worst-case scenario is using QIDS-16, where the probability of a favourable response is 62% with a 95% credible interval of 42%-82% and the best-case scenario is using the MADRS measure, where the probability of a favourable response is 82% with a 95% credible interval of 66%-96%. See Figure 1 for the posterior distribution of the treatment success for both datasets, all data and code associated can be found at https://github.com/szb37/Bayesian-RWE.

Using Bayesian analysis of real-world data, we showed with data from just 20 patients that the probability of success treating childhood epilepsy with cannabis is 95%. In comparison, a traditional placebo-controlled RCT of a very similar product, Epydiolex that contains purified cannabidiol (CBD), in a similar clinical population used a cohort of 170 patients to reveal a statistically significant between-treatment difference using NHST (Thiele et al., 2018). It is worth contrasting the differences between these approaches investigating cannabis's efficacy.

The most obvious one is the difference in the sample size. Placebo-controlled RCTs are 'information inefficient' because half of the patients are not receiving the treatment under investigation. This contributes to the high-cost of drug development that is eventually passed onto the consumers. Moreover, placebo-controlled RCTs raise ethical concerns when an ineffective treatment as in the placebo arm could have severe consequences, e.g. suicide in depression. Traditional RCTs generally require a larger sample, because their primary outcome is the 'between-treatment difference', i.e. how much better is treatment relative to placebo, which requires that about half of the patients are randomized to the placebo group. However, this 'between-treatment difference' is not relevant to either patient or doctors when choosing a treatment. Patients and doctors experience / observe, and care about, the 'change over time', i.e. how much improvement is to be expected from the treatment. Relatedly, traditional RCTs report the between-treatment p-value, which is the '*probability of obtaining the observed or more extreme data assuming that the treatment is no better than placebo*'. Note this probability is not related to treatment success. In contrast the Bayesian analysis yields the probability of treatment success when treating the next patient, which is what doctors and patients care about.

The main reason why RCTs are held in such high regard is because after blinding non-specific treatment effects should equally distributed between treatment arms (Colagiuri, 2010), hence, the between-treatment difference should corresponds to the true treatment effect, free of subjective biases. This is a compelling reason to favour RCTs over alternatives; however, in practice only a small minority of trials measure blinding integrity and thus empirically demonstrate that patients were genuinely unaware of their treatment allocation (Baethge et al., 2013; Fergusson et al., 2004). In many trials participants unblind due to side effects (Colagiuri et al., 2019; Scott et al., 2022), undermining the purpose of blinding and hence the objectivity of placebo-controlled RCTs (Szigeti et al., 2022). In particular, psychedelics elicit conspicuous subjective effects that make them easy to distinguish from placebo, therefore, running truly blinded trials is near impossible (Muthukumaraswamy et al., 2021). In a recent trial of psilocybin for the treatment of alcohol use disorder, 94% of participants correctly guessed their treatment allocation (50% would be expected in a truly blind trial) with a mean confidence of 89% (Bogenschutz et al., 2022). We emphasize that the unblinding is not due to incompetence, but rather the nature of the treatment. Similarly, most exercise / meditation / diet based therapies are tested in a unblinded manner (Knapen et al., 2015).

One major criticism of real-world evidence is the lack of a control condition, raising the question whether the effects could be driven by placebo response. While this concern cannot be entirely eliminated, but there is reason to think there is more at play here. In both of our case series patients were treatment resistant, specifically for the depression study, all patients had failed on at least two previous antidepressant drugs, some on more than ten, and all failed

on psychotherapy. It has been shown that for depression previous failed medication is associated with decreased chances of success for the next treatment (Boswell et al., 2012; Hunter et al., 2015) and the observed response rate is much higher than what's generally observed in the placebo arms of antidepressant trials (~30%) (Walsh et al., 2002), arguing that these results are unexplainable by the placebo response alone. However, control conditions can be incorporated to real-world evidence (RWE). For example, we previously run a 'self-blinding' trial on psychedelic microdosing, where citizen scientists implemented their own placebo control without clinical supervision (Szigeti et al., 2021). An other example of an RWE control condition is provided by a single case where medical cannabis treatment was interrupted due to medicine access problems and seizures rapidly reappeared (Schlag et al., 2022). This case can be viewed as an n=1 ABA(B) trial, i.e. a within-subject crossover trial (Goyal et al., 2022). Therefore, when ethical considerations allow it, such designs could help to establish the causal effect of the treatment even in an RWE context (Barlow & Hersen, 1973; West et al., 2000).

Our arguments here should not be read as a call for the abolition of traditional RCTs, but rather to consider complementary forms of evidence. For example, Bayesian analysis of RWE can be implemented as a hypothesis generating step prior to conducting the more resource intensive traditional RCTs. We believe Sir Michael Rawlins would support this agenda as he argues that *"randomised controlled trials [...] should be replaced by a diversity of approaches that involve analysing the totality of the evidence base"* (Rawlins, 2008).

Acknowledgements

We would like to acknowledge Rayyan Zafar who provided useful feedback.

Conflicts of Interest

B.Sz. and L.P. declares no conflict. D.N. is an advisory to COMPASS Pathways, Neural Therapeutics, and Algernon Pharmaceuticals; received consulting fees from Algernon, H. Lundbeck and Beckley Psytech; received lecture fees from Takeda and Otsuka and Janssen plus owns stock in Alcarelle, Awakn and Psyched Wellness.

References

Baethge, C., Assall, O. P., & Baldessarini, R. J. (2013). Systematic review of blinding assessment in randomized controlled trials in schizophrenia and affective disorders 2000-2010. *Psychotherapy and Psychosomatics*, *82*(3), 152–160. https://doi.org/10.1159/000346144

Barlow, D. H., & Hersen, M. (1973). Single-case experimental designs. Uses in applied clinical research. *Archives of General Psychiatry*, *29*(3), 319–325. https://doi.org/10.1001/archpsyc.1973.04200030017003

Bogenschutz, M. P., Ross, S., Bhatt, S., Baron, T., Forcehimes, A. A., Laska, E., Mennenga, S. E., O'Donnell, K., Owens, L. T., Podrebarac, S., Rotrosen, J., Tonigan, J. S., & Worth, L. (2022). Percentage of Heavy Drinking Days Following Psilocybin-Assisted Psychotherapy vs Placebo in the Treatment of Adult Patients With Alcohol Use Disorder: A Randomized Clinical Trial. *JAMA Psychiatry*. https://doi.org/10.1001/jamapsychiatry.2022.2096

Boswell, J. F., McAleavey, A. A., Castonguay, L. G., Hayes, J. A., & Locke, B. D. (2012). Previous mental health service utilization and change in clients' depressive symptoms. *Journal of Counseling Psychology*, *59*(3), 368–378. https://doi.org/10.1037/a0028078

Carhart-Harris, R. L., Bolstridge, M., Rucker, J., Day, C. M. J., Erritzoe, D., Kaelen, M., Bloomfield, M., Rickard, J. A., Forbes, B., Feilding, A., Taylor, D., Pilling, S., Curran, V. H., & Nutt, D. J. (2016). Psilocybin with psychological support for treatment-resistant depression: An open-label feasibility study. *The Lancet. Psychiatry*, *3*(7), 619–627. https://doi.org/10.1016/S2215-0366(16)30065-7

Colagiuri, B. (2010). Participant expectancies in double-blind randomized placebo-controlled trials: Potential limitations to trial validity. *Clinical Trials (London, England)*, *7*(3), 246–255. https://doi.org/10.1177/1740774510367916

Colagiuri, B., Sharpe, L., & Scott, A. (2019). The Blind Leading the Not-So-Blind: A Meta-Analysis of Blinding in Pharmacological Trials for Chronic Pain. *The Journal of Pain*, *20*(5), 489–500. https://doi.org/10.1016/j.jpain.2018.09.002

Fergusson, D., Glass, K. C., Waring, D., & Shapiro, S. (2004). Turning a blind eye: The success of blinding reported in a random sample of randomised, placebo controlled trials. *BMJ : British Medical Journal*, *328*(7437), 432. https://doi.org/10.1136/bmj.37952.631667.EE

Goyal, P., Safford, M. M., Hilmer, S. N., Steinman, M. A., Matlock, D. D., Maurer, M. S., Lachs, M. S., & Kronish, I. M. (2022). N-of-1 trials to facilitate evidence-based deprescribing: Rationale and case study. *British Journal of Clinical Pharmacology*, *88*(10), 4460–4473. https://doi.org/10.1111/bcp.15442

Hunter, A. M., Cook, I. A., Tartter, M., Sharma, S. K., Disse, G. D., & Leuchter, A. F. (2015). Antidepressant treatment history and drug-placebo separation in a placebo-controlled trial in major depressive disorder. *Psychopharmacology*, *232*(20), 3833–3840. https://doi.org/10.1007/s00213-015-4047-2

Knapen, J., Vancampfort, D., Moriën, Y., & Marchal, Y. (2015). Exercise therapy improves both mental and physical health in patients with major depression. *Disability and Rehabilitation*, *37*(16), 1490–1495. https://doi.org/10.3109/09638288.2014.972579

Kruschke, J. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*. Academic Press.

Muthukumaraswamy, S. D., Forsyth, A., & Lumley, T. (2021). Blinding and expectancy confounds in psychedelic randomized controlled trials. *Expert Review of Clinical Pharmacology*, *14*(9), 1133–1152. https://doi.org/10.1080/17512433.2021.1933434

Rawlins, M. (2008). De testimonio: On the evidence for decisions about the use of therapeutic interventions. *The Lancet*, *372*(9656), 2152–2161. https://doi.org/10.1016/S0140-6736(08)61930-3

Schlag, A. K., Zafar, R. R., Lynskey, M. T., Athanasiou-Fragkouli, A., Phillips, L. D., & Nutt, D. J. (2022). The value of real world evidence: The case of medical cannabis. *Frontiers in Psychiatry*, *13*. https://www.frontiersin.org/journals/psychiatry

Scott, A. J., Sharpe, L., & Colagiuri, B. (2022). A systematic review and meta-analysis of the success of blinding in antidepressant RCTs. *Psychiatry Research*, *307*, 114297. https://doi.org/10.1016/j.psychres.2021.114297

Similon, M. von M., Paasche, C., Krol, F., Lerer, B., Goodwin, G. M., Berk, M., Meyer-Lindenberg, A., Ketter, T. A., Yatham, L. N., Goldberg, J. F., Malhi, G. S., El-Mallakh, R., Licht, R. W., Young, A. H., Kapczinski, F., Swartz, M., Hagin, M., Torrent, C., Serretti, A., … Popovic, D. (2022). Expert consensus recommendations on the use of randomized clinical trials for drug approval in psychiatry- comparing trial designs. *European Neuropsychopharmacology*, *60*, 91–99. https://doi.org/10.1016/j.euroneuro.2022.05.002

Szigeti, B., Kartner, L., Blemings, A., Rosas, F., Feilding, A., Nutt, D. J., Carhart-Harris, R. L., & Erritzoe, D. (2021). Self-blinding citizen science to explore psychedelic microdosing. *ELife*, *10*, e62878. https://doi.org/10.7554/eLife.62878

Szigeti, B., Nutt, D., Carhart-Harris, R., & Erritzoe, D. (2022). *On the fallibility of placebo control and how to address it: A case study in psychedelic microdosing*. PsyArXiv. https://doi.org/10.31234/osf.io/cjfb6

Thiele, E. A., Marsh, E. D., French, J. A., Mazurkiewicz-Beldzinska, M., Benbadis, S. R., Joshi, C., Lyons, P. D., Taylor, A., Roberts, C., Sommerville, K., Gunning, B., Gawlowicz, J., Lisewski, P., Beldzinska, M. M., Szewczyk, K. M., Steinborn, B., Zolnowska, M.,

Hughes, E., McLellan, A., … Wilfong, A. (2018). Cannabidiol in patients with seizures associated with Lennox-Gastaut syndrome (GWPCARE4): A randomised, double-blind, placebo-controlled phase 3 trial. *The Lancet*, *391*(10125), 1085–1096. https://doi.org/10.1016/S0140-6736(18)30136-3

Walsh, B. T., Seidman, S. N., Sysko, R., & Gould, M. (2002). Placebo Response in Studies of Major DepressionVariable, Substantial, and Growing. *JAMA*, *287*(14), 1840–1847. https://doi.org/10.1001/jama.287.14.1840

West, S. G., Biesanz, J. C., & Pitts, S. C. (2000). Causal inference and generalization in field settings: Experimental and quasi-experimental designs. In *Handbook of research methods in social and personality psychology* (pp. 40–84). Cambridge University Press.

Zafar, R., Schlag, A., & Nutt, D. (2020). Ending the pain of children with severe epilepsy? An audit of the impact of medical cannabis in 10 patients. *Drug Science, Policy and Law*, *6*, 205032452097448. https://doi.org/10.1177/2050324520974487

Zafar, R., Schlag, A., Phillips, L., & Nutt, D. J. (2021). Medical cannabis for severe treatment resistant epilepsy in children: A case-series of 10 patients. *BMJ Paediatrics Open*, *5*(1), e001234. https://doi.org/10.1136/bmjpo-2021-001234

Figures

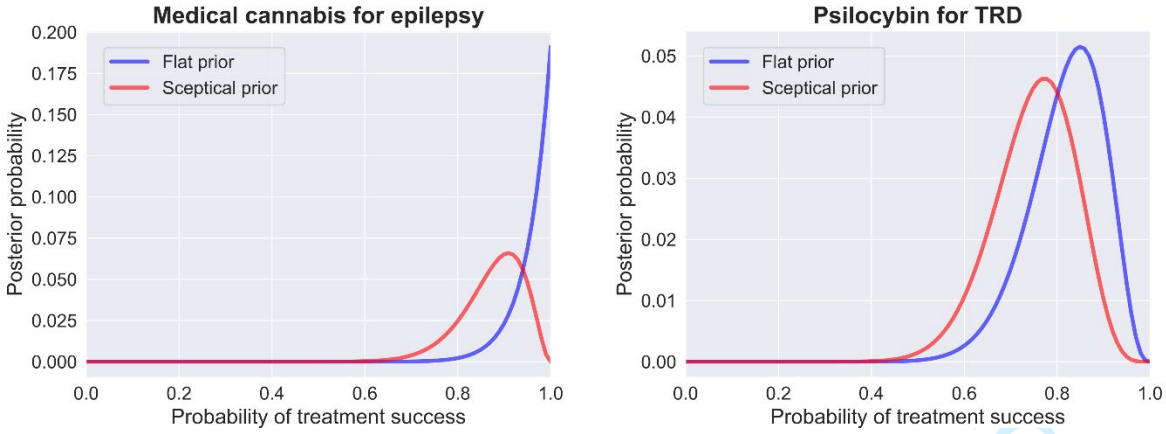

Figure 1: posterior probability of treatment success using the medical cannabis for childhood epilepsy (left) and psilocybin for treatment resistant-depression datasets (right; MADRS outcome measure). Results are shown for priors that are flat (any success rate between and including 0 and 100% is possible) and sceptical (gentle, excluding 0 and 100%, with a mean at 25%).