

BMJ Open Benchmarks for detecting 'breakthroughs' in clinical trials: empirical assessment of the probability of large treatment effects using kernel density estimation

Branko Miladinovic,¹ Ambuj Kumar,^{1,2} Rahul Mhaskar,¹ Benjamin Djulbegovic^{1,2}

To cite: Miladinovic B, Kumar A, Mhaskar R, *et al*. Benchmarks for detecting 'breakthroughs' in clinical trials: empirical assessment of the probability of large treatment effects using kernel density estimation. *BMJ Open* 2014;**4**:e005249. doi:10.1136/bmjopen-2014-005249

► Prepublication history for this paper is available online. To view these files please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2014-005249>).

Received 12 March 2014
Revised 16 September 2014
Accepted 25 September 2014

ABSTRACT

Objective: To understand how often 'breakthroughs,' that is, treatments that significantly improve health outcomes, can be developed.

Design: We applied weighted adaptive kernel density estimation to construct the probability density function for observed treatment effects from five publicly funded cohorts and one privately funded group.

Data Sources: 820 trials involving 1064 comparisons and enrolling 331 004 patients were conducted by five publicly funded cooperative groups. 40 cancer trials involving 50 comparisons and enrolling a total of 19 889 patients were conducted by GlaxoSmithKline.

Results: We calculated that the probability of detecting treatment with large effects is 10% (5–25%), and that the probability of detecting treatment with very large treatment effects is 2% (0.3–10%). Researchers themselves judged that they discovered a new, breakthrough intervention in 16% of trials.

Conclusions: We propose these figures as the benchmarks against which future development of 'breakthrough' treatments should be measured.

INTRODUCTION

Testing of new treatments in randomised controlled trials (RCTs) is widely considered the gold standard for evaluating medical interventions. The process has occurred at an incremental pace and overtime delivered substantial benefits in terms of significant improvement in morbidity and mortality.¹ Ideally, all phases of drug development should be conducted as RCTs² and under the principle of equipoise, which means that no randomised groups should have a predictable treatment advantage over others. When evaluated cumulatively overtime, the probability of detecting a new, successful treatment exerting small or large effect on health outcomes is about 50–60%.¹

Strengths and limitations of this study

- Traditionally, randomised controlled studies have been designed to detect small to moderate differences in treatment effects. At this moment, the benchmarks for 'breakthroughs,' that is, treatments that significantly improve health outcomes against which the discovery rates of new treatment should be measured, are unknown. By providing the overall assessment of the probability of detecting breakthrough interventions, we propose the benchmarks for the future drug development.
- All included trials were of high quality.
- Access to only one industry funded cohort and generalisability of the results to privately funded trials.
- The distribution of observed treatment effects could have been affected by bias in the form of inferior established treatments or other well-documented biases plaguing randomised trials.

Recent developments in molecular techniques and genomic sequencing have led to biomarker driven therapies that often produce large treatment effects ('breakthrough interventions') and the application of novel predictive biomarkers has gained increased importance in the discovery and development of new treatments.^{3–5} RCTs designed with biomarker primary outcomes target specific and usually small patient subgroups, for which the current stringent system of clinical trials that recruits hundreds of patients over the period of many years, may not be optimal.⁶ Therefore, in addition to policy and ethical challenges,⁷ of particular interest is whether this trend will be more effective than RCTs reporting patient relevant primary outcomes in developing new therapies with large treatment effects. Recent published work has suggested that RCTs



CrossMark

For numbered affiliations see end of article.

Correspondence to

Dr Branko Miladinovic;
bmiladin@health.usf.edu

reporting biomarker primary outcomes are more likely to report larger treatment effects than trials reporting patient relevant primary outcomes,⁸ and that RCTs with large treatment effects are more likely than others to have laboratory defined efficacy end points and less likely than others to pertain to death.⁹

How successful these new initiatives will eventually be in the development of new treatments is not known. One way to measure success of RCTs with biomarker primary outcome is to compare it against the existing benchmarks based on patient relevant primary outcomes. The evaluation of research output, such as the estimation of the probability of developing therapies that can substantially improve health outcomes (so-called breakthrough treatments), is of ethical, scientific, and public importance, but has rarely been assessed systematically. We provide such a benchmark based on the analysis of RCTs conducted over half a century across different fields and disciplines by calculating the probability of large treatment effects.

METHODS

Data source

We studied the distribution of treatment effects on mortality and primary outcomes as defined by the original researchers from all consecutive published and unpublished series of completed phase 3 RCTs assessing superiority, registered at onset and conducted by five publicly funded groups (US National Cancer Institute, US National Institute of Neurological Disorders and Stroke, UK Medical Research Council, UK Health Technology Assessment Programme, and National Cancer Institute of Canada Clinical Trials Group) and one privately funded cohort (GlaxoSmithKline). Only the trials that compared new versus established treatments in humans were eligible. The comprehensive list of RCTs, along with associated protocols, was collected by searching the Cochrane Library from the inception to March 2010, Medline Ovid 1950 to March 2010, and EMBASE Ovid 1980 to March 2010. All studies in which the publication bias could not be ruled out were excluded. In the case of unpublished publically funded studies, data were obtained from the cooperative groups. In the case of GlaxoSmithKline, data were extracted from trial summary reports.

The data were collected as part of a Cochrane review,¹⁰ where a detailed description of the search methods, selection criteria, data collection and analysis can be found. The main aims of the review were to: (1) summarise available evidence of the distribution, direction, magnitude and statistical significance of treatment effects; (2) answer the question “What is the probability of new treatments being more effective than established treatments;” (3) test whether the observed distribution of treatment effects is consistent with the ‘uncertainty principle.’ We compared industry versus publically sponsored trials separately.¹¹ The risk of bias and random

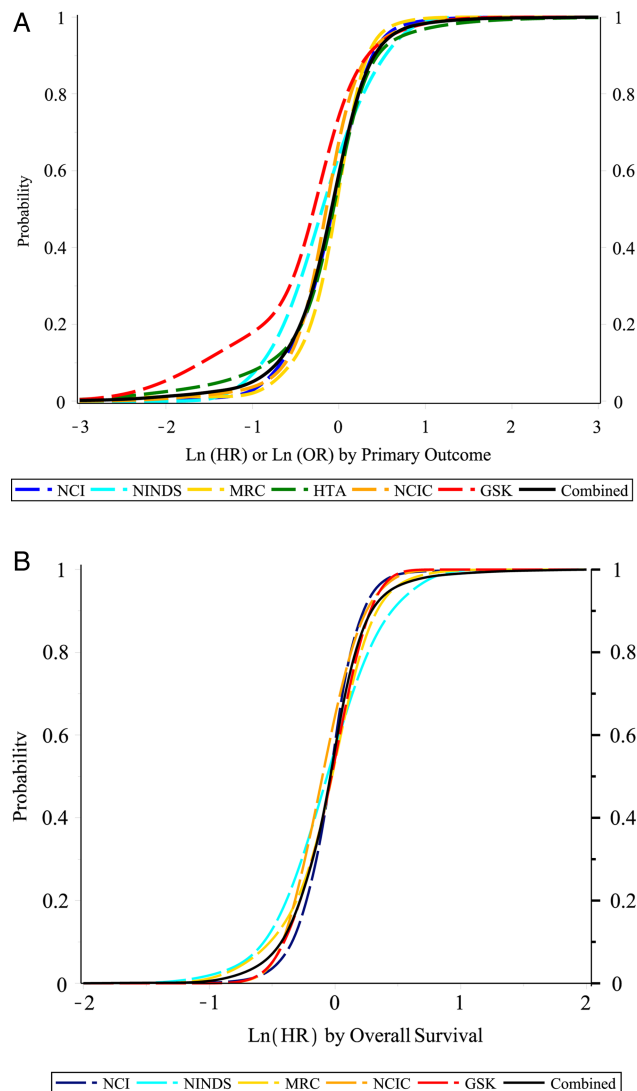


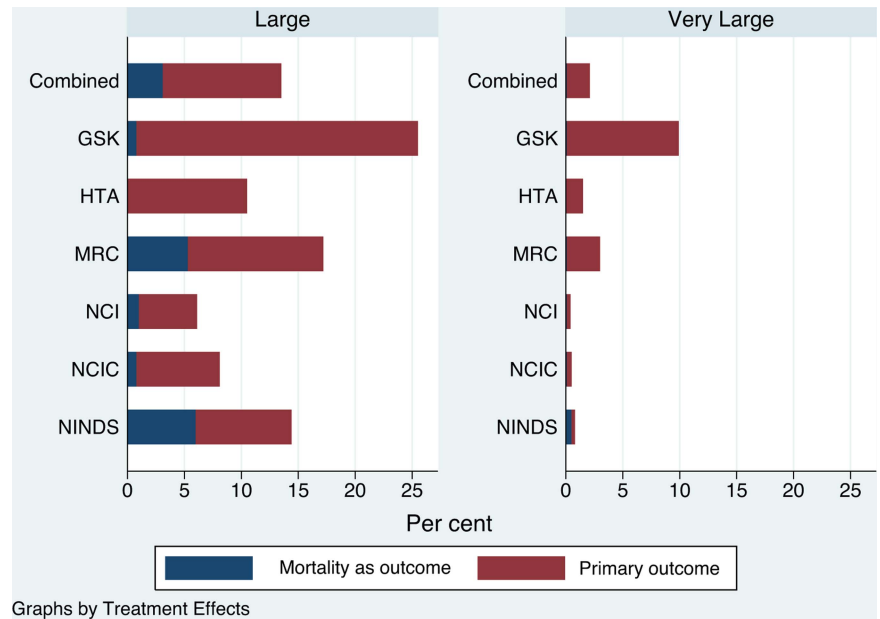
Figure 1 The cumulative probability distributions for all cohorts and combined. NCI, US National Cancer Institute; NINDS, US National Institute of Neurological Disorders and Stroke; MRC, UK Medical Research Council; HTA, UK Health Technology Assessment Programme; NCIC, National Cancer Institute of Canada Clinical Trials Group; GSK, GlaxoSmithKline. HTA did not conduct trials where mortality was an outcome; (A) Cumulative kernel densities for all cohorts using single comparison for each study and weights from random-effects model by Primary Outcome; (B) Cumulative kernel densities for all cohorts using single comparison for each study and weights from random-effects model by overall survival (none of the HTA trials reported overall survival therefore no data were available from this cohort).

error in included studies was assessed systematically according to established methods.¹²

Probability of treatment effects

The probability of large or very large treatment effects was calculated by constructing the empirical distribution of treatment effects using weighted Gaussian kernel

Figure 2 The probability of detecting large or very large treatment effects in RCTs based on the probability distribution of the primary outcome and mortality. The probability distribution was constructed using kernel density estimation.¹⁰ NCI, US National Cancer Institute; NINDS, US National Institute of Neurological Disorders and Stroke; MRC, UK Medical Research Council; HTA, UK Health Technology Assessment Programme; NCIC, National Cancer Institute of Canada Clinical Trials Group; GSK, GlaxoSmithKline.



density estimation,^{13–15} which are based on smoothing histogram given a predefined bandwidth and with the potential of giving different weights to each trial.¹⁴ The probability density function for the hazard or odds ratio on the natural log scale was constructed using two-stage adaptive weighted kernel density estimation. Similar to random effects meta-analysis, the weights were calculated as the inverse of the within-study variance of treatment effects plus the between-study variance for all trials.¹⁰ The estimation was performed using the computation software Maple (V.16). Computational details are further described in the online supplementary appendix.

Size of treatment effects

It is commonly accepted that large treatment effects are those associated with a twofold reduction in relative risk (RR) ($0.2 < RR < 0.5$), and that very large treatment effects are those associated with a fivefold reduction in relative risk ($RR < 0.2$).¹⁶ We also report the per cent of trials where the intervention was so successful according to judgments of the original researchers who thought the newly discovered 'breakthrough treatment' should be adopted immediately as a new therapeutic standard.

RESULTS

Eight hundred and twenty trials involving 1064 comparisons enrolling 331 004 patients were conducted by publicly funded cooperative groups¹⁰ and GlaxoSmithKline conducted 40 cancer trials involving 50 comparisons enrolling a total of 19 889 patients.¹¹ All cohorts included high-quality RCTs with low risk for bias.¹⁰ Figure 1A, B show the cumulative kernel density estimation of the effects of new treatments compared to the established ones for both primary outcomes and overall survival. The analysis according to primary outcomes is

considered important as it reflects the original design and the trialists' 'best bets' that new treatments may prove to be superior to established ones¹⁰ while the analysis according to overall survival relates to pooling data on most important outcomes for patients.

We calculated (figure 2) that in terms of the effects of new treatments on primary outcomes, on average researchers detected the large treatment effects with the probability of 10% (5–25%), or very large treatment effects at 2% (0.3–10%). In terms of the effects of new treatments on mortality, on average, researchers detected the large treatment effects with the probability of 3% (0.8–5.3%), or very large treatment effects at 0.1% (0.05–0.5%). Additionally, in 16% of all trials the original researchers recommended that the new treatment should be adopted immediately as a new standard of care (15% (n=124) publically versus 35% (n=14) privately funded RCTs, $p < 0.001$).

DISCUSSION

Over the past half century, RCTs have delivered steady advances in identifying ways of improving treatments leading to reduced morbidity and increased life expectancy. They also have at times detected large treatment effects. Traditionally, RCTs have been designed to detect small to moderate differences in treatment effects. Under the auspices of a highly predictive marker that predicts the likely response to a specific treatment, some have argued in favour of transforming the current clinical trial system by relaxing the desired false positive rates⁶ and/or increasing the treatment difference worthy of detection.¹⁷ By providing the overall assessment of the probability of detecting 'breakthrough interventions,' we propose the benchmark for the future drug development. Establishment of a benchmark for

the development of new, breakthrough interventions is important for all interested in the therapeutic discovery process: funders can obtain realistic expectations of new treatment development, researchers can use our data to better design new trials, and patients can be better informed what to expect when they volunteer to participate in RCTs.

Author affiliations

¹Division and Center for Evidence-Based Medicine and Outcomes Research, Department of Internal Medicine, Morsani College of Medicine, University of South Florida, Tampa, Florida, USA

²Department of Hematology and Health Outcomes Research, H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida, USA

Contributors BM wrote the manuscript and performed data analysis. BD designed the study and assisted with the manuscript preparation. AK and RM assisted with the manuscript preparation.

Funding This work has been funded by the grant support from the US National Institute of Health for the projects 'Treatment Success and Ethical Principle of Equipoise' (grant no: 1R01CA140408-01); 'Equipoise and the research integrity of clinical trials' (grant no: 1 R01 NS044417-01); 'Evaluation of the quality of clinical trials', (1 R01 NS052956-01), "Quality of Research on treatment harms in cancer", (1R01CA133594-01) (PI: Dr Djulbegovic).

Competing interests None.

Patient consent Obtained.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

REFERENCES

1. Djulbegovic B, Kumar A, Glasziou P, *et al*. Medical research: trial unpredictability yields predictable therapy gains. *Nature* 2013;500:395–6.
2. Djulbegovic B, Hozo I, Ioannidis JP. Improving the drug development process: more not less randomized trials. *JAMA* 2014;311:355–6.
3. Freidlin B, McShane LM, Korn EL. Randomized clinical trials with biomarkers: design issues. *J Natl Cancer Inst* 2010;102:152–60.
4. Orloff J, Douglas F, Pinheiro J, *et al*. The future of drug development: advancing clinical trial design. *Nat Rev Drug Discov* 2009;8:949–57.
5. McShane LM, Cavenagh MM, Lively TG, *et al*. Criteria for the use of omics-based predictors in clinical trials. *Nature* 2013;502:317–20.
6. Deley MC, Ballman KV, Marandet J, *et al*. Taking the long view: how to design a series of Phase III trials to maximize cumulative therapeutic benefit. *Clin Trials* 2012;9:283–92.
7. Hudson KL. Genomics, health care, and society. *N Engl J Med* 2011;365:1033–41.
8. Ciani O, Buyse M, Garside R, *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:f457.
9. Pereira TV, Horwitz RI, Ioannidis JP. Empirical evaluation of very large treatment effects of medical interventions. *JAMA* 2012;308:1676–84.
10. Djulbegovic B, Kumar A, Glasziou PP, *et al*. New treatments compared to established treatments in randomized trials. *Cochrane Database Syst Rev* 2012;10:MR000024.
11. Djulbegovic B, Kumar A, Miladinovic B, *et al*. Treatment success in cancer: industry compared to publicly sponsored randomized controlled trials. *PLoS ONE* 2013;8:e58711.
12. Higgins JPT, Altman DG, Gotzsche PC, *et al*. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;343:d5928.
13. Rosenblatt M. Remarks on some nonparametric estimates of a density-function. *Ann Math Stat* 1956;27:832–7.
14. Silverman BW. *Density estimation for statistics and data analysis*. London: Chapman and Hall, 1986.
15. Silverman BW. Kernel density-estimation using the fast fourier-transform. *Appl Stat J Royal Stat Soc Ser C* 1982;31:93–9.
16. Guyatt GH, Oxman AD, Sultan S, *et al*. GRADE guidelines: 9. Rating up the quality of evidence. *J Clin Epidemiol* 2011;64:1311–16.
17. Sleijfer S, Bogaerts J, Siu LL. Designing transformative clinical trials in the cancer genome era. *J Clin Oncol* 2013;31:1834–41.