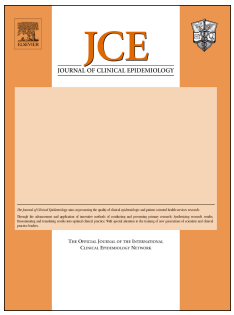
Journal Pre-proof



Theory of Clinical Therapeutic Progress: Reconciling Equipoise With Fat-Tailed (Skewed) Outcomes

Iztok Hozo, PhD, Benjamin Djulbegovic, MD, PhD

PII: S0895-4356(25)00337-3

DOI: https://doi.org/10.1016/j.jclinepi.2025.112004

Reference: JCE 112004

To appear in: *Journal of Clinical Epidemiology*

Received Date: 14 July 2025

Revised Date: 20 September 2025

Accepted Date: 24 September 2025

Please cite this article as: Hozo I, Djulbegovic B, Theory of Clinical Therapeutic Progress: Reconciling Equipoise With Fat-Tailed (Skewed) Outcomes, *Journal of Clinical Epidemiology* (2025), doi: https://doi.org/10.1016/j.jclinepi.2025.112004.

Theory of Clinical Therapeutic Progress: Reconciling Equipoise With Fat-Tailed (Skewed) Outcomes

Running Title:

Connecting equipoise and fat-tailed outcomes to guide therapeutic advances

Iztok Hozo, PhD[1] & Benjamin Djulbegovic, MD, PhD[2]

[1]Department of Mathematics, Indiana University Northwest, Gary, Indiana, USA

[2]Medical University of South Carolina, Division of Medical Hematology and Oncology, Department of Medicine, Charleston, SC, USA

Word count:
Abstract: 325
Plain Language Summary: 179
Manuscript: 3,000

Correspondence:

Benjamin Djulbegovic, MD, PhD
Professor of Medicine
Medical University of South Carolina
39 Sabin Street
MSC 635
Charleston, SC 29425

e-mail: djulbegov@musc.edu
Phone: 843-792-4879

Abstract [325]

Objective

Randomized controlled trials (RCTs) embody a paradox: they require equipoise—genuine uncertainty about treatment effects—to ethically protect participants, yet this very requirement can slow therapeutic innovation by discouraging trials of promising new treatments. We sought to determine which statistical distribution best reflects the balance between protecting trial participants and advancing therapeutic discoveries.

Study Design and Setting

We performed a systematic review to analyze 716 cancer RCTs (1955–2018; published by 2022) encompassing 984 experimental versus standard treatments and ~350,000 patients. Treatment effects were expressed as odds ratios (ORs >1 favoring new therapy). We compared multiple statistical models to identify the distribution that best describes observed effects while preserving equipoise.

Results

Treatment effects were not normally distributed but followed a piecewise log-normal–generalized Pareto distribution (log-normal–GPD). This model captured the heavy right tail of large treatment effects (~3% breakthroughs) better than competing distributions. Entropy, a measure of uncertainty, reached 96%—a modest 4% reduction from the theoretical maximum under normal-theory maximum but sufficient to maintain near-maximum unpredictability in patient-level randomization. Importantly, the heavier tail of the log-normal–GPD model increased the probability of identifying breakthrough therapies by ~3% without undermining the ethical principle of 50:50 allocation. Standard assumptions of normality, frequently used in meta-analysis, overstated precision and failed to capture extreme outcomes, whereas the log-normal–GPD more accurately reflected trial data.

Conclusion

We show that linking ethics (equipoise) with the science of uncertainty (entropy) identifies the optimal statistical distribution driving therapeutic discoveries (log-normal–GPD). Our model preserves ethical unpredictability for patients while modestly enhancing the likelihood of breakthroughs, thereby reconciling patient protection with societal need for innovation. Our findings highlight the limitations of assuming normality in RCTs and meta-analyses and call for reevaluating statistical assumptions in trial design, Bayesian prior specification, and guideline development. Recognizing the skewed distribution of treatment effects can accelerate therapeutic progress while maintaining the ethical foundation of clinical research. Fat tails enable breakthroughs; thin tails truncate and hide them.

Plain Language Summary [179]

Medical progress relies on research that ultimately must rest, in part, on experimentation in [randomized] clinical trials (RCTs) involving human subjects. Researchers are ethically required to protect trial patients from being knowingly exposed to inferior treatments. This is achieved by requiring that investigators are truly uncertain whether new treatments are better than others. On the other hand, funders and investigators spend years developing new treatments. As a result, they often have informative hunches about which treatments would be better. Under which circumstances do the societal interests in developing new treatments align with the protection of patients volunteering in trials? We analyzed a large number of cancer trials published from 1955 to 2022 and found that a special statistical approach (log-normal–GPD distribution) best supports both developing better treatments while upholding patient protection in RCTs. Adopting these principles in future research could enhance the conduct of clinical trials, leading to advances in therapeutics and improved methods for developing treatment guidelines, ultimately benefiting both society at large and individual patients. When ethics align with science, the discovery of new treatments advances.

The development of new therapeutic agents lies at the intersection of the science of drug development and the ethics of human experimentation.[1,2] Randomized controlled trials (RCTs)- an "indispensable ordeal"[3] - are indisputably the most superior method for assessing the effects of health interventions provided that both investigators and patients maintain a genuine state of uncertainty—commonly referred to as equipoise[ftnote1]— regarding the benefits and harms of competing treatment alternatives. Ethicists have generally agreed that requiring uncertainty about treatment choices is essential for protecting trial patients by ensuring they have a fair chance of receiving the optimal intervention before the trial begins.[1] If outcomes were predictable and treatment allocations known in advance, patients might be knowingly exposed to inferior treatments, and research would neither resolve key uncertainties nor use resources effectively. [4] Although treatments may eventually prove superior or inferior to standard care, this can only be determined *after* trial completion.[1] In fact, the unpredictability of individual trial outcomes is what underpins the overall progress in therapeutic discoveries. [1,5-7] Importantly, there is a predictable relationship between uncertainty before trials are initiated and after they are completed. [1,5-7]It is precisely due to the unpredictability of the results in individual trials that average "predictable therapy gains"[7] in the discovery of new treatments have been realized.

According to the equipoise hypothesis, if investigators are truly uncertain about the effects of competing treatment alternatives, a normal statistical distribution over the aggregate of the trials would theoretically maximize entropy [8] — offering the fairest chance of assigning patients to either the experimental or standard treatment, and predicting new treatments as superior 50% of the time.[1,5]

[1] [The general requirement for uncertainty as a condition for the trial has never been challenged by ethicists; the debate has focused on the question "whose uncertainty is morally relevant?"—uncertainties of individual physicians (individual equipoise), patients and physicians (uncertainty principle), community of experts (community equipoise), or individual patients (indifference principle).[1] Here, we reserve the notion of equipoise as the specific measure of uncertainty relating to the information-theoretical concept of entropy, as further discussed in the paper]

However, our previous studies show that treatment successes do not follow a normal distribution but instead display a skewed pattern that favors new treatments. [6,9 10-13]

Observations of this asymmetry suggest that a model of equipoise focusing solely on patient interests is unrealistic. Clinical trials rely not only on patient participation but also on the investments of researchers, funders, and society. Since it is impractical to test every idea, only treatments that appear sufficiently promising advance to RCTs. Although investigators are bound by equipoise, their informed "bets" inevitably increase the probability of success.[14,15] The fact that new treatments are more often successful than standard ones raise concerns that the trial system may not fully uphold the ethical uncertainty required to protect patients. This trade-off introduces an inherent tension: preserving maximum uncertainty while optimizing the discovery of effective treatments.

To date, no theoretical or empirical assessments have identified the statistical distribution that best balances unpredictability in treatment effects with consistent therapeutic advances. Understanding this distribution is crucial to advancing clinical progress and ensuring the rational allocation of resources.[16,17] In this paper, we propose that both theoretical considerations and empirical data support a piecewise lognormal–Generalized Pareto Distribution (logNormal–GPD) as the optimal model. This distribution not only explains observed outcomes but also serves as a platform for accelerating therapeutic progress.

Methods

We consider a typical two-group RCT where new treatment is compared with standard therapy to assess their effects on outcomes expressed on the continuum. Although our methods apply to all human trials, we illustrate our approach using cancer studies, where the most extensive data on therapeutic progress has been collected.

We have assembled cohorts of RCTs meeting the following criteria to minimize publication bias and include unpublished data: (i) a consecutive series of trials, (ii) registration at or before study onset, and

(iii) comparison of new treatments with established ones in humans.[18] The database comprises five cohorts of RCTs evaluating various supportive and disease-oriented treatments over the past 65 years across different cancers. [6 19 20 21] It includes 716 trials comparing 984 experimental treatments with standard treatments in 349,947 patients, conducted between 1955 and 2018 and published by 2022. Generally, these were well-conducted studies with no evidence of bias, although many used placebo or no treatment as the comparator.[13] We assessed treatment effects based on primary outcomes (e.g., response rates, progression-free survival, overall survival) as defined in each trial. Details on cohort's provenance, time window, tumor types, endpoints, and comparator classes were previously reported separately [6 19 20 21] and jointly. [9,13] [see also Supplement1(S1), PRISMA diagram and Checklist].[22]

We expressed treatment effects as hazard ratios (HRs) or odds ratios (ORs) (hereafter denoted as OR in the analysis section), harmonizing the data so that OR > 1 indicates improvement in desirable outcomes. To our knowledge, no other datasets in the literature have such a long track record of evaluating new versus standard treatments.

Theoretical assumptions

Theoretical distributions of equipoise can take many forms.[23] As stated earlier, the empirical data on treatment successes did not adhere to normal distribution [6] unless the tails, which capture large effects, are ignored. We previously determined that the tails of successful, new treatments adhere to (generalized) Pareto distribution [9,13] as expected according to the Pickands–Balkema–de Haan theorem.[24]

The observed asymmetry in distribution of treatment effects does not necessarily breach the uncertainty requirement. When patients are randomized to experimental versus standard treatments, it remains unpredictable whether effects will fall within the main body or the tails of the distribution.

However, investigators typically pursue “positive” signals, creating selection bias: thousands of molecules are screened, but only promising candidates enter trials. This tilts research toward optimism and larger

effects[25]; thus, RCT priors are enriched with the skewed priors. Since only treatments with promising early results advance to RCTs, perfect symmetry between experimental and standard outcomes is not expected. Instead, uncertainty is maintained in the main body, while a heavy tail increases the likelihood of success for new treatments—a pattern reflecting the mechanism of preferential attachment, where success begets further success.[26-28] Thus, to fully understand a relationship between uncertainty and therapeutic advances, the most important task is to determine to which theoretical distribution the empirical data adhere. Knowing which distribution data follows is crucial for choosing the right analysis.

Among several candidate models, the piecewise log-normal Generalized Pareto Distribution (log-normal–GPD) is the most plausible. It preserves the unpredictability of the log-normal body [patient-level uncertainty] while adding extra uncertainty in the tail [researchers' informed expectations].

We therefore set out to assess: 1) which statistical distribution best fits our empirical data, and 2) which model maintains ethical equipoise by preserving high unpredictability while allowing a higher chance of breakthroughs. For context, we compare our findings against other distributions, commonly assumed models in medical research[29] (Table 1).

Because our aim is to depict the full empirical distribution of treatment effects—including both central and extreme values—we did not apply shrinkage or regularization methods to obtain a representative estimate of the average treatment effect when outliers or noise are present. Indeed, shrinking large effects toward the mean would mask truly exceptional treatment effects. However, because research synthesis (meta-analysis) of clinical trials is essential for evidence-based medicine and the development of guidelines [30], we conducted a meta-analysis using kernel density, distribution-free methods [18] and compared the results with standard meta-analyses that assume treatment effects are normally distributed.

Calculation of uncertainty and fitting statistical distribution to empirical data

We use entropy as a measure of uncertainty. [8,31,32] First, we calculated the entropy of each statistical distribution fitted to the observed treatment effects. In RCTs, we are concerned with the uncertainty in choosing between treatments. We use information theory to define maximum entropy—when all outcomes are equally likely—as "equipoise".[8,31,32] In other words, equipoise reflects an even (symmetric) distribution of uncertainty about the relative merits of competing treatments, serving as a precise measure of maximal uncertainty. [8,31,32] In practice, equipoise is typically expressed qualitatively through researchers' hunches or predictions, [4,33] accepting randomization "only in our state of ignorance". [34]

Consequently, we related entropy to treatment success when the effect size, expressed as an odds ratio (OR), favors the new experimental treatment (OR > 1) ("success") over the standard treatment (OR < 1) ("failure"). Specifically, entropy is defined as:

$$H(p) = -[p\log_2(p) + (1-p)\log_2(1-p)] \qquad \text{(eq 1)}$$

where p = Probability (OR>1) according to the given statistical distribution and H is entropy expressed in bits. The equation shows that when the likelihood of success and failure is equal [$\Pr(OR \geq 1) \rightarrow p = 0.5$], the entropy reaches its maximum value of 1.

Although there is no generally accepted definition of breakthrough treatments effects, the treatments that exceed threshold effects of relative risk or OR = 2, 5, 3, 10, or 12 often are considered breakthroughs.[9,35-37] Because these categorizations are based on the OR thresholds, we fitted piecewise log-normal-GPD model to our empirical data. Using maximum likelihood estimation, we modeled the body of the distribution with a log-normal function and the tails with a GPD (see S2 for details).

For our default analysis, we set a threshold at T=OR=3 to secure a reliable estimate in the tails, where large effects concentrate. We defined small effects as $1<OR<1.2$, moderate: $1.2 \leq OR<3$ and large effects OR>3. This approach allowed us to calculate the probability of small, moderate vs. large treatment effects, and subsequently relate it to entropy- the main goal of this analysis.

In addition to the threshold (T), GPD is characterized by two other parameters: Scale parameter (denoted by σ), which is analogous to the role of standard deviation in normal distributions, and the shape parameter (denoted by ξ), which governs tail behavior. A positive ξ indicates no upper bound on the maximum effect detectable in trials.[13]

To validate our model fit, we used Q-Q plots to compare the observed density of treatment effects with theoretical predictions and applied the Kolmogorov-Smirnov (K-S) test to assess the null hypothesis of no difference between the fitted and observed data. A non-significant K-S test and the straight line on the Q-Q plots indicate a good fit. Sensitivity analyses confirmed our estimates were stable, with optimal results at T = 3 and maximum OR = 20. All analyses were implemented in Stata statistical package.[38] [Data with statistical codes are provided in the Supplementary Material.]

Thus, our approach was based not only on empirical analysis of treatment-effect distributions but also on theoretical considerations of equipoise and entropy, incorporating both patient and researcher perspectives on anticipated treatment effects. This approach was postulated a priori and corroborated through statistical analyses.

Results

Figure 1 displays the fitted statistical distributions at T = 3. Both log-normal and log-normal–GPD models visually fit well, with about 3% of data clustering in the heavy right tail.

Figure 2 presents K-S tests and Q-Q plots comparing the empirical data to the two most competitive theoretical models. While the K-S tests were not significant, the Q-Q plots favor the log-normal–GPD model, which better captures the heavy tail than the log-normal distribution. Other distributions performed poorly (see also S3).

Fig 3 shows the probability of treatment success. As theoretically predicted, probability of treatment success [favoring new treatment, OR>1 = probability of small effects (1<OR<1.2) + moderate effects (1.2<OR<3) + large effects (OR>3)] and the probability of treatment failure (favoring standard treatment, OR<1) is close to 50%:50% according to the hypothetical normal (benchmark) distribution.

However, our empirical data show a 63:37% split favoring new treatments. For other distributions, the success rates were as follows: 65% (normal), 58% (log-normal), 61% (log-normal–GPD), 43% (exponential), 63% (scaled beta), and 60% (Weibull). Notably, only the log-normal–GPD model provided a satisfactory fit and predicted ~3% large effects ("breakthroughs"), closely matching the observed 3.05%. Most treatment successes were moderate in size. (Fig 1, S4 Table; Fig S5)

S4 Table summarizes the findings and regression analysis. The shape parameter (ξ) was greater than 0, suggesting no upper bound on detectable effects, although confidence intervals permitted $\xi < 0$; observed treatment effects ranged from OR 3 to 20.

Figure 4 shows that, while a hypothetical normal distribution reaches 100% entropy for binary choices, entropy for all other models varies narrowly between 93% and 98%. The log-normal–GPD model, the only valid model, is associated with 96% entropy and about a 3% chance of detecting large effects. Fig S5 depicts cumulative distributions of all models. Fig S6 shows a high correlation between entropy and the probability of detecting large treatment effects (Spearman $\rho = 0.8183$, $p = 0.0301$).

Fig 5. Comparison of meta-analysis using the standard random-effects Sidik-Jonkman model, which assumes treatment effects follow a normal distribution, versus kernel-density, distribution-free meta-analysis, which accounts for skewed distributions of treatment effects. Note the false sense of certainty and precision produced by the standard model—uncertainties that become apparent when using empirically more appropriate, skewed distributions.

Discussion

Our findings show that treatment success follows a log-normal–GPD distribution, with entropy dropping from 100% to 96% while yielding a 3% chance of breakthrough (large) effects. This reduction does not violate equipoise; it preserves near-maximal unpredictability at the individual patient level while enhancing the discovery of meaningful treatments. In this way, the trade-off supports both ethical patient protection and the motivation of researchers and funders. The log-normal–GPD model thus aligns ethical considerations with the practical societal needs of drug development, guaranteeing 'predictable therapy gains' [7] while maintaining necessary uncertainty. It satisfies the key precepts of moral philosophy and helps link the theory of human experimentation with the theory of rational choice.[39]

Interestingly, studies indicate that the general public accepts RCTs with equipoise as skewed as 70:30[40], and IRB members approve them when the imbalance is no greater than 80:20.[41,42] Moreover, refusal rates to participate in a RCT (at 50:50 random allocation) increase when the probability of a new treatment's superiority exceeds 80%. [41,42]

Our results have major implications for evidence interpretation and conduct of RCTs. Standard meta-analyses typically assume normality or log-normality of treatment effects, assumptions that can fail when data are skewed or heavy-tailed. Models like the log-normal–GPD better capture extreme outcomes and improve uncertainty quantification, reducing the risk of overlooking (of approximately 3%) valuable treatments. Our model joins the calls for a re-evaluation of the statistical assumptions underlying the design of RCTs, meta-analyses, and clinical guidelines. [6,9 10-12,43] As noted long ago "normality is a myth; there never was, and never will be , a normal distribution". [44,45] Additionally, one of the major challenges in the contemporary conduct and analysis of Bayesian trials concerns the choice of prior distributions.[46] Our data suggest that log-normal–GPD distribution can be effectively used as priors in the design of Bayesian randomized controlled trials, finally providing an answer to one of the most important questions in the history of clinical therapeutics—posed by Iain Chalmers in 1997: "What is the

prior probability of a proposed new treatment being superior to established treatments?" [47] Similarly, Liu at el[10] reported that between 15% and 26% of Cochrane reviews violated the normality assumption, which is also reflected in our findings (see Fig 5). [A full discussion of optimal methods for meta-analyzing skewed data—is beyond the scope of this paper and warrants further research.]

Our study's primary limitation is its focus on cancer RCTs. However, as long as equipoise is maintained, we expect similar log-normal–GPD patterns in other fields. That is, theoretically, it is possible that outcomes in individual RCTs adhere to a normal distribution, but in aggregate behave according to a log-normal–GPD distribution. However, this is even more likely in the case of cancer because outcomes in individual cancer RCTs are rarely normally distributed.[48] Therefore, it is not surprising that data based on the means and standard deviations of each trial adhere to a non-normal, log-normal–GPD distribution, as demonstrated herein.[49]

We should note that many large effects in our database came from small, supportive-care trials tested against placebo, yet numerous interventions discovered in these studies have remained in practice for over 30 years, supporting the validity of their effects.[6,13,18] Nevertheless, even if some trials are biased, our analysis indicates that the requirement for uncertainty has not been violated—at least at the macro level—to the extent that the entire system of RCTs would be considered unethical.

We also used outcome distributions as a proxy for priors since it was not possible to elicit anticipated treatment effects from trials spanning 65 years. Prior research confirms that observed outcomes across many trials are not arbitrary but predictable[6,7,18,19,50-52], constrained by prior beliefs. The decision to randomize embeds the prior beliefs of patients (via consent/ethics) and researchers (via design/funding).

Finally, although this work continues two decades of pre-registered research [6,18], this analysis was not pre-registered.

Conclusions

We linked the principle of equipoise with entropy-based uncertainty metrics to identify the (fat-tailed) statistical distributions that best support the discovery of new treatments RCTs. We showed that the log-normal–GPD model balances uncertainty with ethics in RCTs to drive therapeutic progress. The log-normal–GPD distribution maintains fair treatment allocation by preserving high uncertainty at the individual level, while its small loss in entropy—coupled with an increased chance of breakthroughs (~3%)—does not compromise ethical randomization. A 50:50 randomization remains ethically acceptable, ensuring that patients are still enrolled in an environment of genuine uncertainty.

In the final analysis, advances in therapeutics are owed to the *paradox of equipoise*[5,7]—while individual trial outcomes are unpredictable, the overall distribution of treatment success follows a predictable log-normal–GPD pattern across many trials- that sustains clinical trials system and drives therapeutic innovation as anengine of therapeutic progress: entropy + (patients' and researchers') equipoise → Optimal RCT design→ aggregate distribution adheres to log-normal + GPD. We argue that the non-normal distribution of trial outcomes has profound implications for how clinical trials are designed, analyzed, aggregated, and interpreted. Because log-normal–GPD fat tails reveal rare breakthroughs while thin-tailed distributions truncate high values, limiting discovery, more treatments could be identified if trial methods evolved to handle non-normal outcomes.

Legends:

Fig 1. Distribution of treatment success in cancer randomized trials published between 1955 and 2022. The fit of several statistical distributions is shown; the best fit was observed for the log-normal–generalized Pareto distribution (log-normal–GPD).
Abbreviations: PDF – probability density function; OR – odds ratio (see also Table 1).

Fig 2. Q–Q (quantile–quantile) plot comparing the distribution of quantiles from empirical data ("real data") against: a) theoretical log-normal–generalized Pareto distribution, and b) theoretical log-normal distribution. The straight lines and non-significant K–S (Kolmogorov–Smirnov) test results indicate a good fit.

Fig 3. Probability of effect sizes observed in empirical ("real") data and according to selected statistical distributions. P(OR < 1) indicates the probability (in percentages) that the standard treatment is superior to the experimental treatment. P(OR > 1) indicates the probability (in percentages) that the experimental treatment is superior to the standard treatment. [Note: P(OR > 1) = small effects (1 < OR < 1.2) + moderate effects (1.2 < OR < 3) + large effects (OR > 3).]

Fig 4. Relationship between overall entropy and the probability of treatment effects based on selected statistical distributions. Maximum entropy is 100% (see Equation 1), which corresponds to a hypothetical normal distribution.

Fig 5. Comparison of meta-analysis using the standard random-effects Sidik-Jonkman model, which assumes treatment effects follow a normal distribution, versus kernel-density, distribution-free meta-analysis, which accounts for skewed distributions of treatment effects. Note the false and inappropriate sense of certainty and precision produced by the standard model—uncertainties that become apparent when using empirically more appropriate, skewed distributions. Kernel-density analysis clearly shows that the global distribution of effects (posterior ensemble across many trials) is asymmetric, consisting of trials demonstrating many 'null/modest' effects, some favoring standard and some experimental treatments, with a few breakthroughs displaying large benefit. [Note that before performing kernel-density meta-analysis [18], we tested the normality assumption in our dataset using the Wang & Lee method[53] to assess the standardized treatment effect, resulting in a Shapiro-Wilk test: $p < 0.001$—indicating a highly significant deviation from normality.]

Acknowledgement:

Datasets for this work were obtained with support of grants from the US National Institute of Health: R01CA140408, R01NS044417, R01NS052956, and R01CA133594 (Djulbegovic).

Contribution: IH created the statistical code, helped revise the manuscript and performed analyses. BD- conceived the idea for the project, wrote the first draft, and performed analyses. Both authors agreed with the final version of the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** This is a reanalysis of data already available in the public domains. Stata files used in this analysis will be made publicly available upon the paper's publication (see also Supplementary Material).

We also want to thank Dr. Amy Price for her useful feedback, which helped improve readability for a general, non-specialist audience.

References

1. Djulbegovic B. Articulating and responding to uncertainties in clinical research. *J Med Philosophy*. 2007;32:79-98.
2. World Medical Association. World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Participants. *JAMA*. 2024;doi:10.1001/jama.2024.21972
3. Frederickson DS. Sorting out the doctor's bag. . *Control Clin Trials* 1980;1:263-7.
4. Djulbegovic B, Hozo I, Greenland S. Uncertainty in Clinical Medicine. In: Gifford F, ed. *Philosophy of Medicine (Handbook of the Philosophy of Science)*. Elsevier; 2011:299-356.
5. Djulbegovic B. The paradox of equipoise: the principle that drives and limits therapeutic discoveries in clinical research. *Cancer Control*. Oct 2009;16(4):342-7.
6. Djulbegovic B, Kumar A, Soares HP, et al. Treatment success in cancer: new cancer treatment successes identified in phase 3 randomized controlled trials conducted by the National Cancer Institute-sponsored cooperative oncology groups, 1955 to 2006. *Arch Intern Med*. Mar 24 2008;168(6):632-42.
7. Djulbegovic B, Kumar A, Glasziou P, Miladinovic B, Chalmers I. Medical research: Trial unpredictability yields predictable therapy gains. . *Nature*. 2013;500(7463):395-396. doi:10.1038/500395a
8. Cover TM, Thomas, Joy A. *Elements of Information Theory*. Wiley. ; 1991.
9. Hozo I, Djulbegovic B, Parish AJ, Ioannidis JPA. Identification of threshold for large (dramatic) effects that would obviate randomized trials is not possible. *Journal of Clinical Epidemiology*. 2022/05/01/ 2022;145:101-111. doi:https://doi.org/10.1016/j.jclinepi.2022.01.016
10. Liu Z, Al Amer FM, Xiao M, et al. The normality assumption on between-study random effects was questionable in a considerable number of Cochrane meta-analyses. *BMC Medicine*. 2023/03/29 2023;21(1):112. doi:10.1186/s12916-023-02823-9
11. Burton C. Heavy Tailed Distributions of Effect Sizes in Systematic Reviews of Complex Interventions. *PLOS ONE*. 2012;7(3):e34222. doi:10.1371/journal.pone.0034222
12. Noma H, Nagashima K, Kato S, Teramukai S, Furukawa TA. Meta-analysis Using Flexible Random-effects Distribution Models. *Journal of Epidemiology*. 2022;32(10):441-448. doi:10.2188/jea.JE20200376
13. Djulbegovic B, Hozo I, Iskander R, Parish AJ, Kimmelman J, Ioannidis JPA. There Is No Upper Limit on the Maximum Effect That Can Be Detected in Randomized Trials. *J Clin Epidemiol*. May 12 2025:111828. doi:10.1016/j.jclinepi.2025.111828
14. Fries JF, Krishnan E. Equipoise, design bias, and randomized controlled trials: the elusive ethics of new drug development. *Arthritis Res Ther*. 2004;6(3):R250-5. doi:10.1186/ar1170

ar1170 [pii]
15. Djulbegovic B, Hozo I. When is it rational to participate in a clinical trial? a game theory approach incorporating trust, regret and guilt. *BMC Medical Research Methodology*. 2012;12(1):85.
16. Brophy JM, Gelman A. Rethinking approaches to analysis of global randomised controlled trials. *BMJ*. 2025;389:r1273. doi:10.1136/bmj.r1273
17. Greenland S. Invited Commentary: The Need for Cognitive Science in Methodology. *American Journal of Epidemiology*. 2017;186(6):639-645. doi:10.1093/aje/kwx259
18. Djulbegovic B, Kumar A, Glasziou PP, et al. New treatments compared to established treatments in randomized trials. Research Support, N.I.H., Extramural

Research Support, Non-U.S. Gov't. *Cochrane Database Syst Rev*. 2012;10:MR000024. doi:10.1002/14651858.MR000024.pub3
19. Djulbegovic B, Kumar A, Miladinovic B, et al. Treatment success in cancer: industry compared to publicly sponsored randomized controlled trials. *PLoS One*. 2013;8(3):e58711. doi:10.1371/journal.pone.0058711

20. Machin D, Stenning S, Parmar M, et al. Thirty years of medical reseach council randomized trials in solid tumours. *Clin Oncology*. 1997;9:100-114.
21. Iskander R, Moyer H, Fergusson D, McGrath S, Benedetti A, Kimmelman J. The Benefits and Risks of Receiving Investigational Solid Tumor Drugs in Randomized Trials : A Systematic Review and Meta-analysis. *Ann Intern Med*. Jun 2024;177(6):759-767. doi:10.7326/M23-2515
22. Liberati A, Altman D, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *J Clin Epidemiol*. 2009;62:e1 - e34.
23. London AJ, Kane PB, Kimmelman J. Varieties of Community Uncertainty and Clinical Equipoise. *Kennedy Inst Ethics J*. 2023;33(1):1-19. doi:10.1353/ken.2023.a899457
24. Coles S. *An Introduction to Statistical Modeling of Extreme Values*. Springer-Verlag; 2001.
25. Djulbegovic B, Kumar A, Magazin A, et al. Optimism bias leads to inconclusive results-an empirical study. *J Clin Epidemiol*. Jun 2011;64(6):583-93. doi:S0895-4356(10)00317-3 [pii]

10.1016/j.jclinepi.2010.09.007
26. Clauset AR, Shalizi R, Newman MEJ. Power-law distributions in empirical data. *SIAM Review*. 2009;51:661-703.
27. Newman MEJ. Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*. 2005;46:323-351.
28. Mahony H, Tsalatsanis A, Kumar A, Djulbegovic B. Evolution of Treatment Regimens in Multiple Myeloma: A Social Network Analysis. *PloS one*. 2014;9(8):e104555.
29. Kimmelman J, London AJ. Predicting Harms and Benefits in Translational Trials: Ethics, Evidence, and Uncertainty. *PLoS Med*. 2011;8(3):e1001010. doi:10.1371/journal.pmed.1001010
30. Djulbegovic B, Guyatt GH. Progress in evidence-based medicine: a quarter century on. *Lancet*. Jul 22 2017;390(10092):415-423. doi:10.1016/S0140-6736(16)31592-6
31. Tsalatsanis A, Hozo I, Djulbegovic B. Research synthesis of information theory measures of uncertainty: Meta-analysis of entropy and mutual information of diagnostic tests. *Journal of Evaluation in Clinical Practice*. 2021;27(2):246-255. doi:https://doi.org/10.1111/jep.13475
32. Weaver W. Some recent contributions to the mathematical theory of communication. *Shannon CE, Waever W The mathematical theory of communication* The University of Illinois Press,; 1962.
33. Wittgenstein L. *On Certainty*. Harper Torchbooks; 1969.
34. Hill AB. The clinical trial. *N Engl J Med*. 1952;(247):113-119.
35. Djulbegovic B, Glasziou P, Klocksieben FA, et al. Larger effect sizes in nonrandomized studies are associated with higher rates of EMA licensing approval. *Journal of Clinical Epidemiology*. 2018;98:24-32. doi:10.1016/j.jclinepi.2018.01.011
36. Djulbegovic B, Glasziou P, Chalmers I. The importance of randomised vs non-randomised trials. *The Lancet*. 2019/08/24/ 2019;394(10199):634-635. doi:https://doi.org/10.1016/S0140-6736(19)31128-6
37. 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-analysisDrug and Device Approvals Based on Nonrandomized Clinical TrialsDrug and Device Approvals Based on Nonrandomized Clinical Trials. *JAMA Network Open*. 2019;2(9):e1911111-e1911111. doi:10.1001/jamanetworkopen.2019.11111
38. *STATA, ver. 18*. 2022.
39. Rawls J. *A theory of justice. Revised edition.* Harvard University Press; 1999.
40. Johnson N, Lilford JR, Brazier W. At what level of collective equipoise does a clinical trial become ethical? *J Med Ethics*. 1991;17:30-34.

41. Mhaskar R, Bercu B, Djulbegovic B. At what level of collective equipoise does a randomized clinical trial become ethical for the members of institutional review board/ethical committees? *Acta Inform Med*. 2013;21(3):156-9.
42. Mhaskar R, Miladinovic B, Guterbock TM, Djulbegovic B. When are clinical trials beneficial for study patients and future patients? A factorial vignette-based survey of institutional review board members. *BMJ Open*. 2016;6(9):e011150. doi:10.1136/bmjopen-2016-011150
43. Jackson D, White IR. When should meta-analysis avoid making hidden normality assumptions? *Biometrical Journal*. 2018;60(6):1040-1058. doi:https://doi.org/10.1002/bimj.201800071
44. Geary RC. Testing for normality. *Biometrika*. 1947;34(Pt 3-4):209-42.
45. Micceri T. The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*. 1989;105(1):156-166. doi:10.1037/0033-2909.105.1.156
46. Lee JJ, Yin G. Principles and Reporting of Bayesian Trials. *J Thorac Oncol*. Jan 2021;16(1):30-36. doi:10.1016/j.jtho.2020.10.010
47. Chalmers I. What is the prior probability of a proposed new treatment being superior to established treatments? *BMJ*. 1997;314:74-75.
48. Collett D. *Modelling Survival Data in Medical Research (3rd ed.).* Chapman and Hall/CRC. ; 2014.
49. Djulbegovic B. What is the probability that results of meta-analyses will favor innovative treatments? presented at: 8th Annual Cochrane Colloqium - 'Evidence for action: challenges for The Cochrane Collaboration in the 21st century'; 2000; Cape Town, South Africa. Accessed August 31, 2022. https://abstracts.cochrane.org/2000-cape-town/what-probability-results-meta-analyses-will-favor-innovative-treatments
50. Kumar A, Soares H, Wells R, et al. Are experimental treatments for cancer in children superior to established treatments? Observational study of randomised controlled trials by the Children's Oncology Group. *Bmj*. Dec 3 2005;331(7528):1295.
51. Soares HP, Kumar A, Daniels S, et al. Evaluation of New Treatments in Radiation Oncology: Are They Better Than Standard Treatments? *JAMA*. February 23, 2005 2005;293(8):970-978.
52. Djulbegovic B, Lacevic M, Cantor A, et al. The uncertainty principle and industry-sponsored research. *Lancet*. 2000;356:635-38.
53. Wang C-C, Lee W-C. Evaluation of the Normality Assumption in Meta-Analyses. *American Journal of Epidemiology*. 2019;189(3):235-242. doi:10.1093/aje/kwz261

Table 1: Comparison of Statistical Distributions for Modeling Treatment Effects: theoretical considerations

Distribution	Range/Support	Entropy Behavior	Tail Behavior (Large Effects)	Suitability for Trials
Normal	(-∞, ∞)	Maximizes entropy for fixed variance (highest overall entropy under those conditions); however, not constrained to positive values.	Tails decay very fast (light tails); almost zero probability for extreme effects.	A theoretical benchmark; not suitable for outcomes (e.g., odds ratios) that must be positive.
Log-normal	(0, ∞)	Lower overall entropy than Normal due to the positivity constraint; most probability mass is concentrated in the body (where moderate effects are observed).	Moderately heavy tail: may predict small factions of the treatment successes clustered in throughout the tails.	Widely used in clinical research for positive outcomes; realistic for many treatment effects but may underestimate extreme outcomes.
Mixed Log-normal–GPD*	(0, ∞)	Slightly higher overall entropy than pure log-normal (extra uncertainty allocated to the tail).	Heavy tail: assigns additional probability to extreme outcomes (i.e., increasing the chance for detecting large treatment effects clustered within the tails).	Very attractive for trials aiming to detect rare, breakthrough effects while still preserving near-maximum unpredictability (equipoise).
Exponential	[0, ∞)	Maximum entropy for a positive variable with a fixed mean; e.g., H = 1 + ln(mean)	Tail decays exponentially; allows for large outcomes but not as heavy as a Pareto-type tail.	Frequently used in survival analysis as a benchmark; represents the maximum-entropy case for a positive variable with fixed mean.
Weibull (Survival)	[0, ∞)	Entropy depends on both shape and scale; very flexible – when shape = 1 it equals the exponential (maximal entropy for	Tail behavior varies with shape: for shape parameter k = 1 it is heavy (exponential),	Widely used in survival analysis; interpretable and flexible for modeling time-to-event data.

Distribution	Range/Support	Entropy Behavior	Tail Behavior (Large Effects)	Suitability for Trials
		a given mean), and for shape > 1 the tail decays faster.	for k > 1 the tail decays faster.	
Binomial (Binary)	{0, 1}	Maximum entropy is 1 bit when p = 0.5; however, it is discrete and provides very limited uncertainty.	Cannot capture gradations in effect; no tail beyond the two outcomes.	Too simplistic for modeling continuous treatment effects; can only be used for binary endpoints; not directly applicable for continuous treatment effect sizes (such as HR or OR)
Beta	[0, 1]	Maximum entropy when uniform; otherwise, entropy decreases as the distribution becomes more peaked.	Bounded support limits tail behavior (cannot exceed 1); unable to represent extremely large outcomes, unless we use a scaling transformation.	Useful for modeling proportions, but not for unbounded outcomes like odds ratios. In general, it can only be used via scaling transformation.
Negative Binomial / Poisson	Discrete counts	Entropy depends on the mean and variance; designed for count data rather than continuous effect sizes.	Tail behavior is fixed by the count model; does not capture the continuous spectrum of extreme outcomes.	Not directly applicable for continuous treatment effect sizes.

*distribution with the best fit to our empirical data (see Fig 1 and 2); OR-odds ratio; HR- hazard ratio

Fig 1.

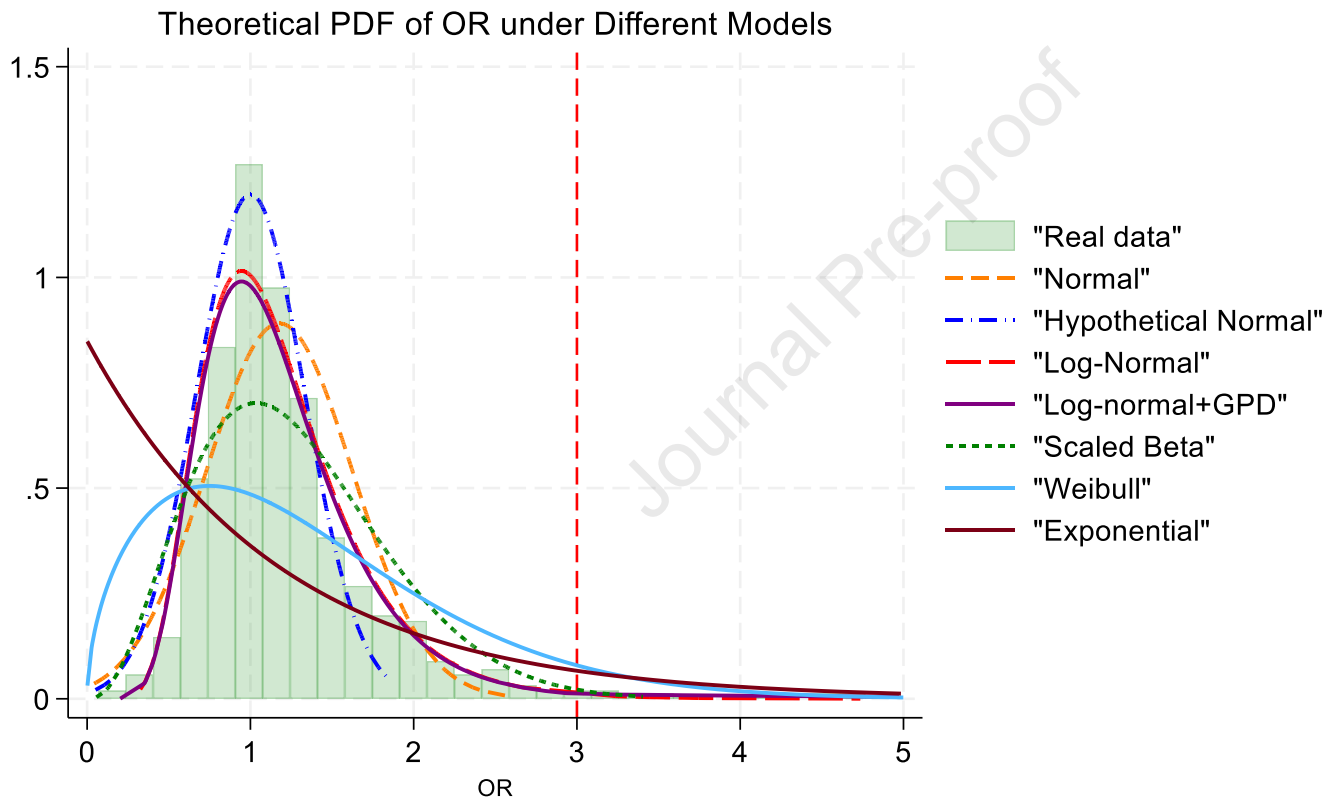
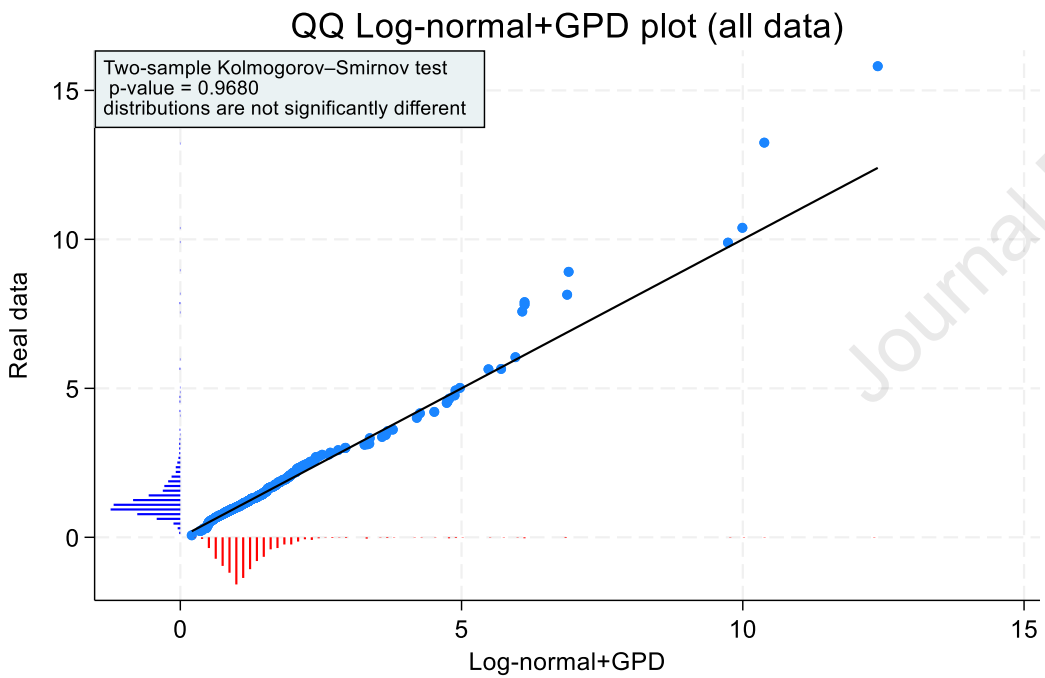


Fig 2.

a)



b)

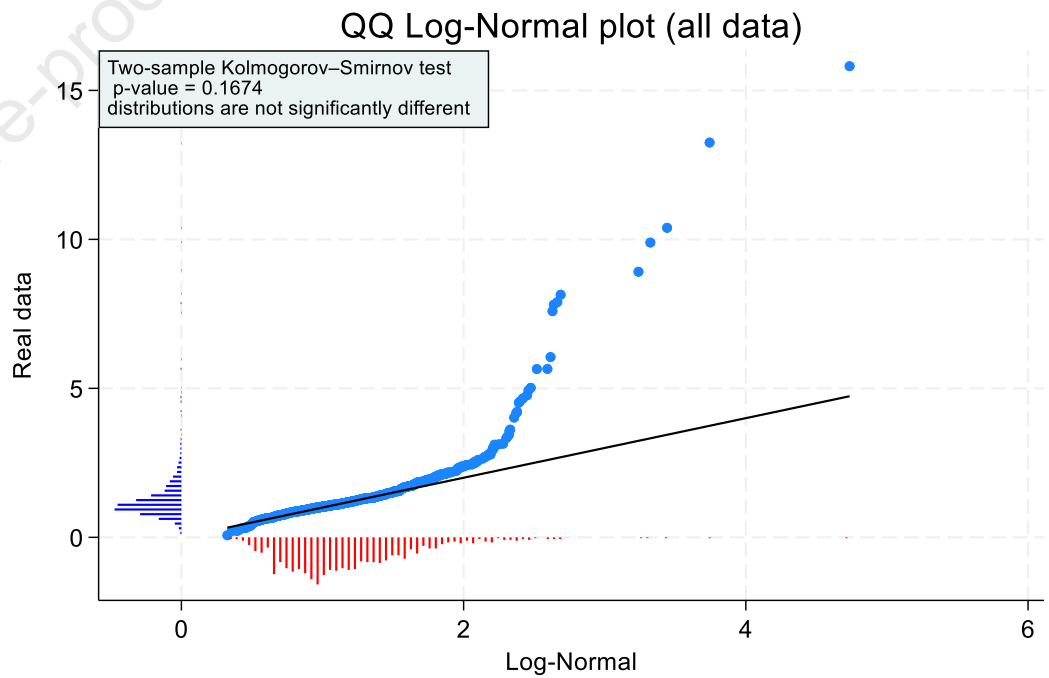


Fig 3.

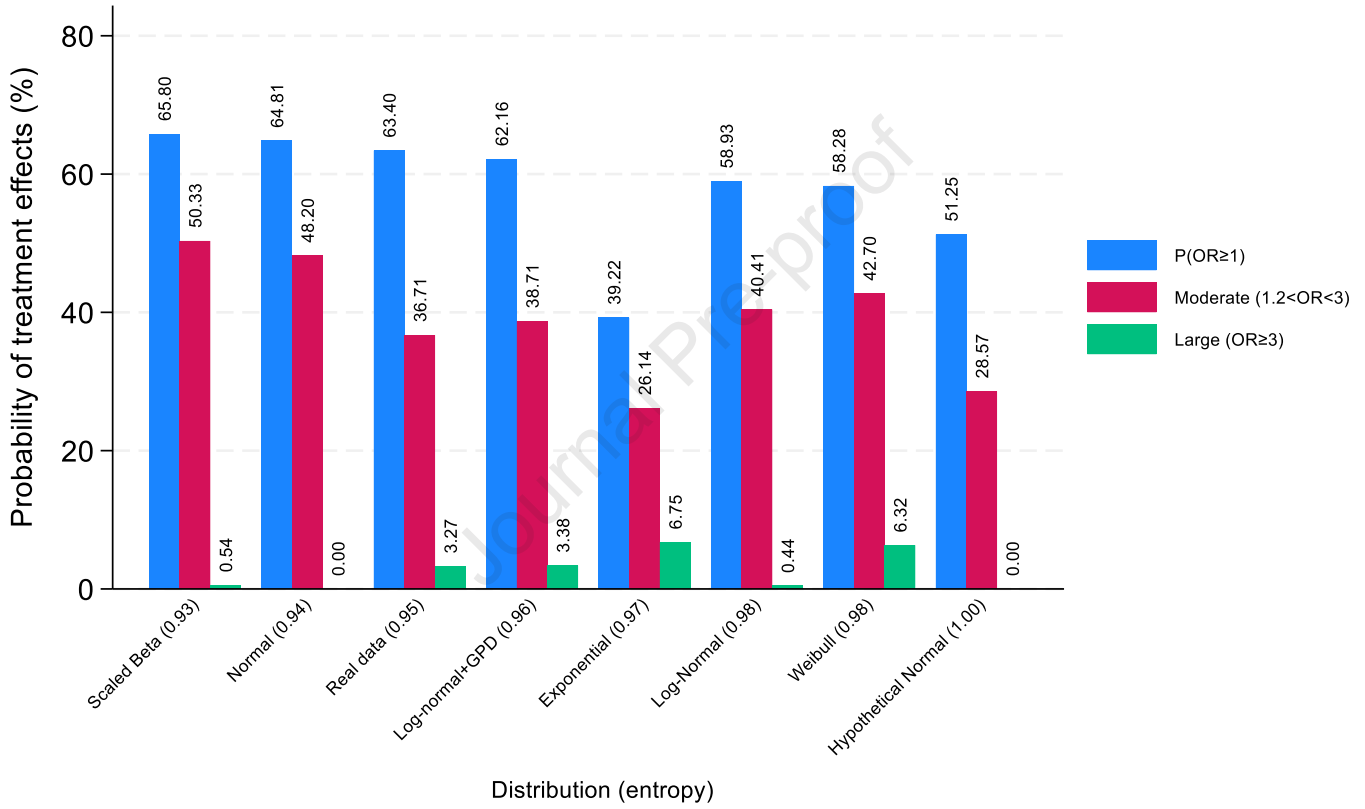


Fig 4.

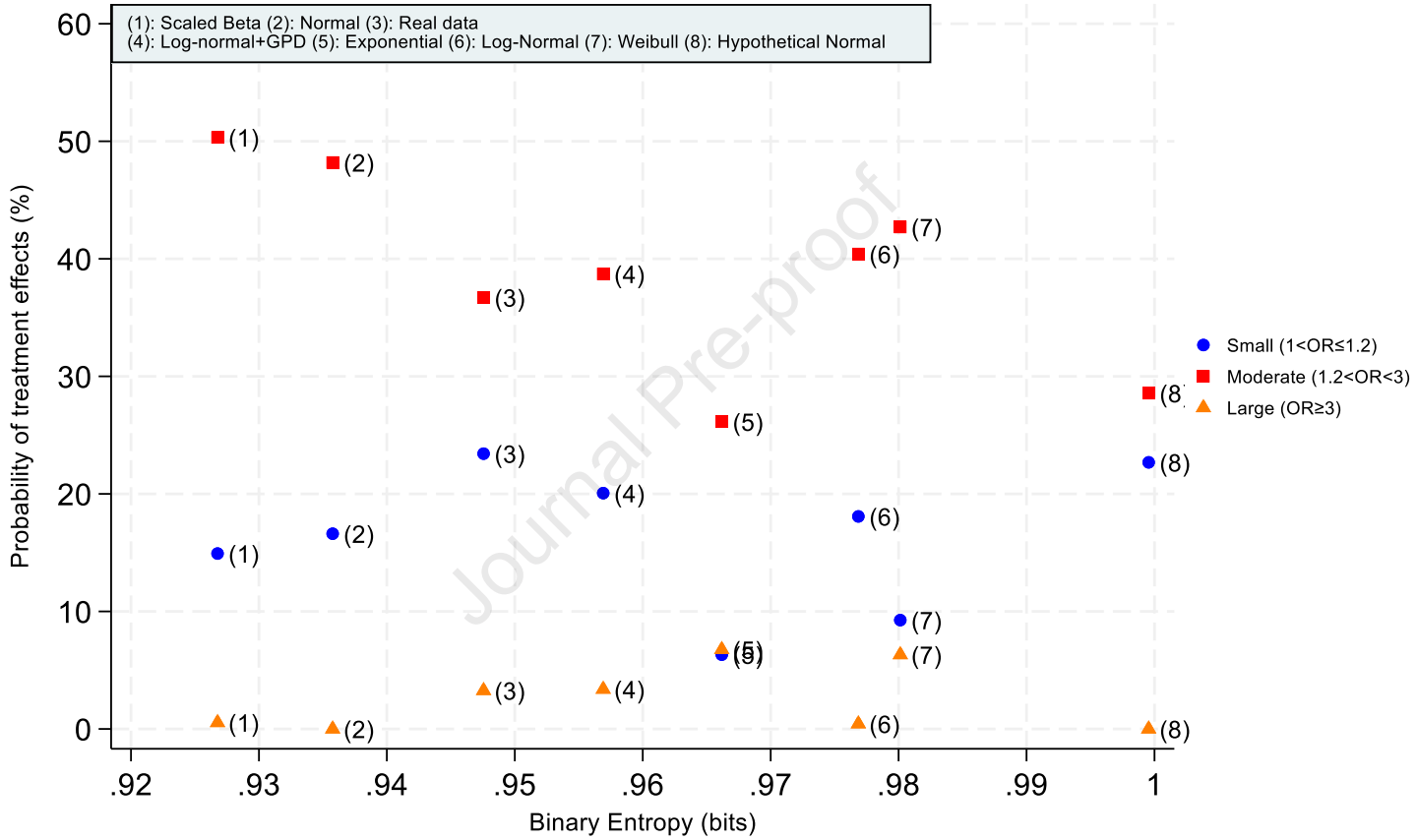
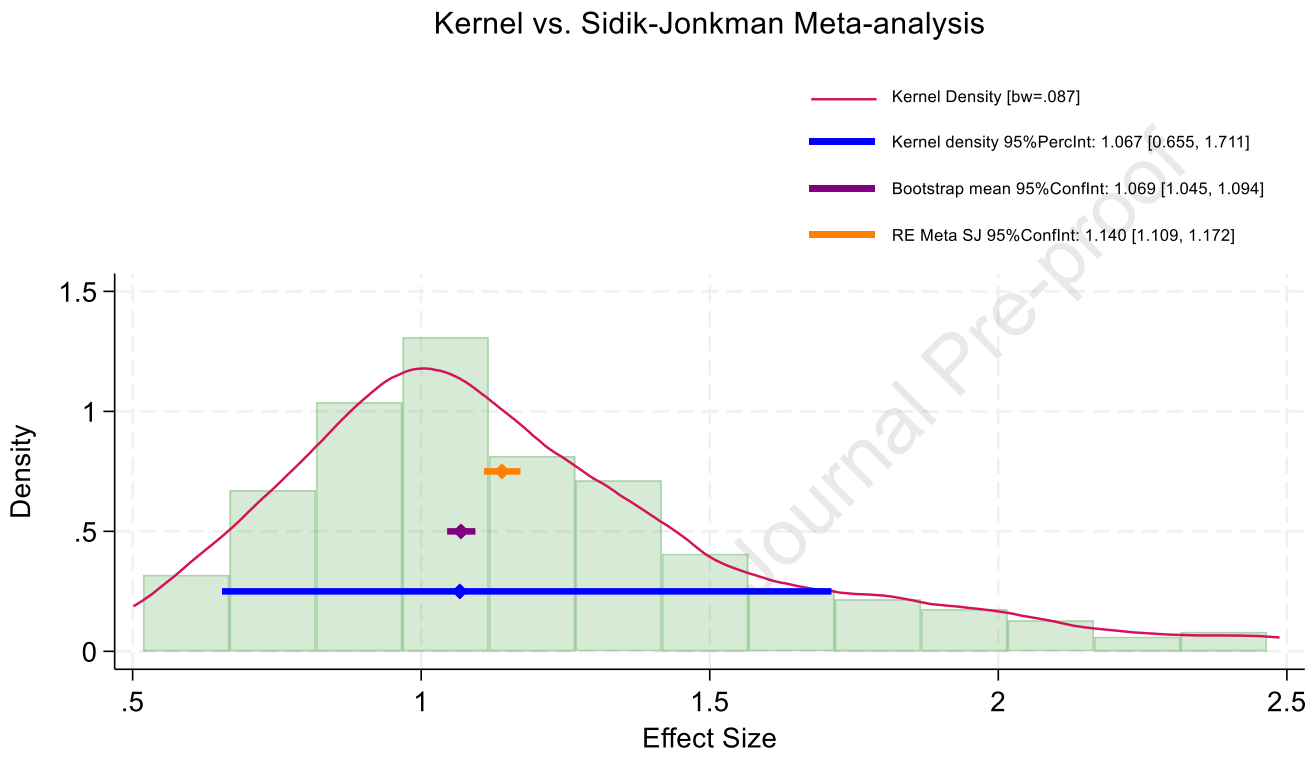


Fig 5.



Highlights

- **RCTs should only be conducted under genuine uncertainty (equipoise)** about the effects of competing treatments; otherwise, results are predictable and randomization is unethical.
- **Equipoise implies a normal distribution of treatment effects,** with maximum uncertainty (entropy) and equal chances of benefit for all patients.
- **In practice, funders and researchers often favor trials in which new treatments are expected to perform better,** shifting the outcome distribution away from normality.
- **Considering both patient and researcher perspectives, treatment effects in RCTs are best modeled by a piecewise, log-normal–GPD distribution,** resulting in only a minor (4%) drop in entropy and preserving ethical equipoise.
- **Recognizing this slightly skewed distribution has major implications for trial design, analysis, systematic reviews, and clinical guidelines,** and could help identify (3-4%) breakthrough treatments that standard assumptions might overlook.

What is known?

- To ensure randomization does not disadvantage patients volunteering for RCTs, it is widely accepted that **trials should only be conducted when equipoise exists**—that is, when the results cannot be predicted in advance.
- **The equipoise hypothesis predicts that treatment outcomes in RCTs should follow a normal distribution**. However, empirical data show that this is often not the case.
- "**Normality is a myth**; there never was, and never will be , a normal distribution."

What is new?

- We show that **treatment effects in RCTs follow a piecewise log-normal–GPD distribution**—a slightly skewed pattern that preserves equipoise while facilitating the discovery of new treatments (3 - 4%) with large effects.

What should change now?

- Our findings have important implications for how clinical research is conducted and interpreted.
- **Most immediately, meta-analyses of RCTs should assume a skewed distribution of treatment effects**.
- Log-normal–GPD distribution can be effectively **used as priors** in the design of Bayesian randomized controlled trials

- Designing RCTs based on **the fat-tailed (log-normal–GPD) assumption may help discover new treatments** that might have gone unnoticed under traditional (normal or symmetric) distribution assumptions.
- **Distributions with fat tails enable breakthroughs; thin tails statistical distributions truncate and hide them.**

Contribution: IH created the statistical code, helped revise the manuscript and performed analyses. BD-conceived the idea for the project, wrote the first draft, and performed analyses. Both authors agreed with the final version of the manuscript.

Declaration of Interest Statement

We declare no conflict of interest – intellectual or financial – in relation to this work.