Review Article

A holistic framework for interpreting positive clinical outcomes in biotechnology

Nikil Varma

From Northeastern University, San Francisco, California, United States

ABSTRACT

Clinical trials conducted by biotechnology companies are vital to not only the success of their products, but the health and safety of future patients. However, positive study results do not always lead to FDA authorization and subsequent release to the market for numerous reasons including, false positive rates, validity and reliability of outcomes, the significance of results, and the semantics of company communications. Driven by the need to compete within the marketplace, biotechnology companies have motivation to distort and/or present results to make them appear more favorable. As a result, it is important to approach positive clinical trial results with a healthy amount of skepticism and utilize a more holistic approach to evaluating results and determining whether the advertised product should be released into the market.

Key words: *Biotechnology, Clinical trials, False positive rate, FDA, Fragility index*

In the realm of biotechnology clinical studies, favorable outcomes in trials do not assure subsequent FDA authorization. In fact, it is crucial to approach positive findings from drug trials with a healthy degree of doubt and to scrutinize various aspects before accepting the results [1]. Certain factors need to be considered when interpreting positive outcomes in biotech clinical including, intrinsic false positive rates (FPR) and hypothesis believability in clinical trials, strength and reliability of trial outcomes, the significance and relevance of trial results, and the semantics of press releases and investor presentations.

INTRINSIC FPRS AND HYPOTHESIS BELIEVABILITY IN CLINICAL TRIALS

In analyzing clinical trials, it is widely acknowledged that reproducing many clinical results can be challenging due to the intrinsic FPR present in every clinical study, regardless of the drug's effectiveness [2]. A key aspect of FPR is the inherent probability of a scientific hypothesis being valid. For instance, before examining a clinical trial investigating the efficacy of Oscillococcinum, a homeopathic remedy derived from duck heart and liver for influenza treatment, one might assume that any positive outcome is likely a false positive; conversely, a comparable study on imatinib, an approved and effective drug for chronic myeloid leukemia, would likely exhibit a very low FPR, with most clinical trials falling between these two extremes [3,4].

The connection between the inherent credibility of a hypothesis and the FPR can be demonstrated in a study with a binary primary endpoint, such as cured or not cured. Given a trial with standard statistical parameters of power (1-b) of 80% and a standard significance level of 0.05, a drug believed to have a 9:1 chance of being effective would have an FPR of 0.7% (Fig. 1). On the other hand, a standard statistical power of 80% and standard significance level of 0.05 with a 1:1 probability of being effective would yield an FPR of 5.9% (Fig. 2).

Accurately calculating the intrinsic odds of a drug's success is challenging. The quantity and quality of supporting data, the extent to which previous preclinical or early clinical studies are believed to predict clinical outcomes in the studied patient population, and an individual's understanding of the disease pathophysiology, drug mechanism of action, and their interrelationship are all aspects taken into account.

In the case of Parkinson's disease, the lack of reliable preclinical models, translational biomarkers, persuasive and accessible genetic data, or well-founded patient selection hypotheses based on underlying biology due to the heterogeneity of the disease all contribute to a historically high probability of many promising first-in-human and proof-of-mechanism trial results being false positives [5]. Likewise, it is challenging to have confidence in a positive clinical trial result for a drug with two or more apparently unrelated phenotypic traits or poorly understood mechanism, even if supported by preclinical and early clinical data, compared to a therapy with a well-defined pharmacological action that aligns with one's comprehension of the disease etiology.

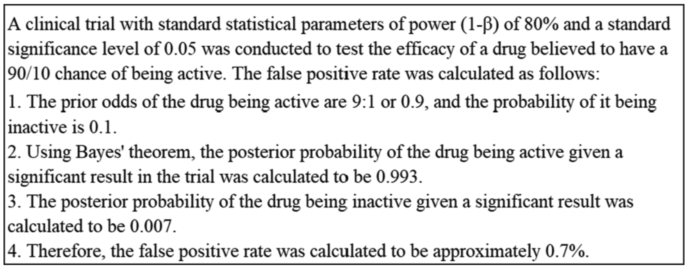
Access this article online

Received - 20 April 2023
Initial Review -24 April 2023
Accepted - 04 May 2023

Quick Response code

DOI: 10.32677/ejms.v8i2.4022

Correspondence to: Nikil Varma, Northeastern University, San Francisco, California, United States. E-mail: varma.ni@northeastern.edu



A clinical trial with standard statistical parameters of power (1-β) of 80% and a standard significance level of 0.05 was conducted to test the efficacy of a drug believed to have a 90/10 chance of being active. The false positive rate was calculated as follows:

1. The prior odds of the drug being active are 9:1 or 0.9, and the probability of it being inactive is 0.1.
2. Using Bayes' theorem, the posterior probability of the drug being active given a significant result in the trial was calculated to be 0.993.
3. The posterior probability of the drug being inactive given a significant result was calculated to be 0.007.
4. Therefore, the false positive rate was calculated to be approximately 0.7%.

Figure 1: Calculation of false positive rate

STRENGTH AND RELIABILITY OF TRIAL OUTCOMES

To assess positive clinical studies, outcome sensitivity needs to be considered. Small positive studies (n<30) are more susceptible to individual responders or non-responders influencing the results. The fragility index (FI) can be used to evaluate the sturdiness of positive trial results, which represents the minimum number of patients whose results must change to switch the overall finding from positive (p<0.05) to negative (p>0.05) [6]. Intuitively, a trial with 30 patients and a p=0.045 is less robust than a similar trial with 1000 patients and the same effect size and P value. FI can be calculated for any study measuring a binary endpoint or an approximate value for time-to-event studies by treating the event as a binary outcome.

The calculation of FI is also applicable to time-to-event outcome-based studies by handling the event-based outcome as a binary and is derived through a process that involves iterative employment of a dichotomous endpoint-appropriate test, such as the Fisher's exact test, to evaluate the impact of modifying patients' outcomes on the P value. While an FI score with a low number of patients does not necessarily imply any trial misconduct or clinical bias, greater confidence in the reproducibility of findings may be warranted if the endpoint measured was overall survival (OS) compared to time-to-progression, as OS is to be a more objective endpoint, as it is defined by a clear, unequivocal event.

To demonstrate the value of FI score, the PROfound Phase 3 randomized, open-label trial evaluated the efficacy and safety of the PARP inhibitor olaparib (Lynparza; AstraZeneca), with enzalutamide (Xtandi; Astellas) or abiraterone (Zytiga; Janssen) as the comparator in patients with mCRPC who have progressed on prior treatment with NHA treatments (abiraterone or enzalutamide) and have a qualifying tumor mutation in BRCA1/2, ATM, or one of 12 other genes involved in the HRR pathway [7]. In 2019, the Phase 3 PROfound trial demonstrated improvement in the primary endpoint of radiographic progression-free survival (rPFS) with a median rPFS of 7.4 months for the active arm versus a median rPFS of 3.5 months for the comparator arm (HR 0.49; p<0.0001) [8]. At 12 months, for Cohorts A and B of the study, approximately 56 of 256 patients in the active arm were progression-free, while 17 of the 131 patients in the control arm were progression-free. Based on the calculated FI (Table 1), if just one incremental patient progressed on olaparib, the P value would become non-significant (p=0.0504) (Table 1). This narrowness in FI demonstrates that one might be more confident in the replicability of clinical results if the endpoints were less subjective.

Another factor to consider in clinical trials is unusual dose-response relationships. An especially alarming situation related to consistency occurs when a clinical trial investigates multiple drug doses but observes significant effects only at low or intermediate doses. Typically, a therapy's effect should increase or stabilize as the dose increases; however, if it declines at higher levels, this may indicate that aspects of the disease or drug are not well understood.

In the BLAZE-1 Phase 2 safety and efficacy randomized, double-blind, and placebo-controlled study that enrolled participants in multiple treatment groups, bamlanivimab (LY-CoV555; Lilly), monoclonal antibody that was specifically designed to target the spike protein of SARS-CoV-2 recently diagnosed outpatients with mild-to-moderate COVID-19. The trial results (PubMed), showed that the middle dose of 2,800 mg had a significant improvement in the primary outcome, which was the change in viral load at day 11 after treatment initiation. However, this significant improvement was not observed at lower or higher doses of 700 mg or 7000 mg. Despite this dose-response relationship, the reason for this idiosyncratic finding remains unknown, as no explanation has been disclosed to date. The development of monotherapy was subsequently discontinued the LY-CoV555 program and shifted to potential combo therapies with bamlanivimab.

SIGNIFICANCE AND RELEVANCE OF POSITIVE RESULTS

In later-phase trials, merely achieving statistical significance might not be adequate to convince regulators, physicians, health-care insurers, or patients of the findings' validity. There are several situations where results with a standard significant level of *P* < 0.05 may be less meaningful than they initially appear. For starters, results deemed statistically significant may not have a clinically significant effect. Since the statistical power can be increased with a larger trial size, statistically significant positive results can be achieved with a weakly effective therapy; a large enough patient population is studied. In 2019, the Phase 3 REGENERATE trial investigated the effects of obeticholic acid (OCA) on liver fibrosis caused by non-alcoholic steatohepatitis (NASH) in patients. The study's primary efficacy analysis revealed that taking OCA 25 mg once a day resulted in an improvement in fibrosis (by at least one stage) without worsening of NASH, meeting the primary endpoint. The planned 18-month interim analysis showed that this improvement was significant compared to taking a placebo (with a P value of 0.0002) [9] However, the FDA informed intercept that the drug's "predicted benefit […] remains uncertain" and does not warrant safety risks in patients with NASH-related liver fibrosis [10]. Specifically, the FDA "felt that the modest fibrosis effect was not clearly predictive of clinical efficacy" [11].

In oncology trials, an inappropriate choice of comparator can lead to some FDA-approved drugs being tested against suboptimal comparators, which can inflate the trial's success odds [12].

Inappropriate choice of irrelevant or poorly predictive endpoint becomes a concern when late-stage trials may use endpoints with

Table 1: Calculated FI

	Active	Comparator
Progression-free	56	17
Progressed	200	114

```
from scipy.stats import fisher_exact
contingency_table = [[56, 17], [200, 114]] |_,
initial_p_value=fisher_exact (contingency_table)
def find_fragility_index (contingency_table, threshold-0.05):
  fragility_index=0
  table = [row.copy() for row in contingency_table]
while True:
  # Reassign one event from the group with more events to the group
  with fewer events max_idx=max (range (2), key-lambda i: table[0]
  [i])
  min_idx=1-max_idx
  table[0][max_idx] = 1
  table[0][min_idx] += 1
  table[1][max_idx] += 1
  table[1][min_idx] -= 1
  # Recalculate the P value
  _p_value=fisher_exact (table)
  # Check if the new P value is no longer statistically significant if
  p_value>threshold: break
  fragility_index+=1
  return fragility_index
fragility_index=find_fragility_index (contingency_table)
Fragility Index: 1
```

	Active	Comparator
Progression-free	55	17
Progressed	201	114

```
from scipy.stats import fisher_exact
updated_contingency_table = [[55, 17], [201, 114]]
updated_p_value=fisher_exact (updated_contingency_table)
Updated p-value: 0.0504
```

little clinical relevance, affecting the drug's perceived efficacy. For example, in rheumatoid arthritis, placebos were used as comparators in 81 out of 102 trials of biologic disease-modifying drugs for rheumatoid arthritis done during the past decade. In 54 (86%) of 63 trials involving patients with a high level of active disease, placebos (or treatments known to have been ineffective) were used, with the result that potentially helpful treatments were being withheld from 9224 out of 13,095 patients randomized to the control arms [12].

Inconsistent effects across key subgroups or endpoints become a concern when clinical data reveals that a drug may be approvable based on a positive primary outcome, but inconsistent, weak, or contradictory findings in key secondary endpoints or subpopulations make it commercially unattractive, especially compared to competitors' products.

Another example can be found in Alzheimer's disease drug trials, where the primary endpoint might be a significant improvement in cognitive function as measured by a standardized test. If secondary endpoints, such as the impact on daily living activities or caregiver burden, show weak or contradictory results in different age groups or disease severity subpopulations, sponsors may discontinue development due to concerns about the drug's commercial viability in the face of competing treatments.

The positive effect of a drug often comes with trade-offs, as the importance of balancing efficacy and safety depends on the specific clinical scenario, but in all cases, it is crucial to assess this balance objectively. For instance, in the study of ramucirumab (Cyramza; Lilly) in NSCLC, the improvement in OS was accompanied by an increased risk of febrile neutropenia, pneumonia, and neutropenia, among other adverse effects [13]. Even in rare diseases with high mortality rates, where patients might be willing to tolerate substantial discomfort and risk for a longer life, there is generally a threshold where the perceived risks and harms outweigh the benefits.

CRITICAL ASSESSMENT OF PRESS RELEASES AND INVESTOR PRESENTATIONS

In many cases, pharmaceutical developers openly report the success or failure of clinical trials; however, there are instances where companies might exaggerate negative or inconclusive outcomes. Such behavior can vary from highlighting positive elements in marginally satisfactory data to blatant deceit intended to hide unfavorable results. Assessing press releases and investor presentations that announce study outcomes can be particularly difficult, as they often omit essential information about the trial's protocol, implementation, and analysis compared to comprehensive journal articles. For example, the press release for the pivotal Phase 3 trial for forigerimod (Lupuzor; ImmuPharma) initially highlighted a "superior response rate over placebo" but further communicated through a press release that results from the pivotal Phase 3 trial were "…not statistically significant" [14,15].

Furthermore, trial sponsors may not clearly communicate a change in a primary endpoint. Therefore, confirming primary endpoints is advisable to verify the primary endpoint – especially when a company releases top-line data. For example, after analysts fretted over Sage Therapeutic's choice of primary endpoint, the sponsor decided to change the endpoint for the Phase 3 trial to changes from baseline in a depression rating scale after three days from a difference in placebo after 15 days [16].

Companies can often identify a subset of patients that appeared to have responded when a trial fails. Subset analyses are susceptible to testing multiple patient groups until a positive result is found. This form of multiple-hypothesis testing is referred to as multiplicity, which is a problem that arises when multiple statistical tests are performed in a single study, increasing the likelihood of finding false positive results by chance alone. Multiplicity should be accounted for in clinical trials through several analysis methods to reduce the risk of false positive findings and maintain the integrity of the study.

REFERENCES

1. Thornquist MD. Understanding the results of clinical trials: A guide for patients and non-statisticians. Cancer Prev Res 2011;4:1193-203.
2. Ioannidis JP. Why most published research findings are false. PLoS Med 2005;2:e124.
3. Mathie RT, Frye J, Fisher P. Homeopathic oscillococcinum® for preventing and treating influenza and influenza-like illness. Cochrane Database Syst Rev 2015;2015:CD001957.
4. Gleevec: The Breakthrough in Cancer Treatment. (n.d.). Nature Education. Available from: https://www.nature.com/scitable/topicpage/gleevec-the-breakthrough-in-cancer-treatment-565 [Last accessed on 2023 Jul 19].
5. Espay AJ, Schwarzschild MA, Tanner CM, *et al*. Biomarker-driven phenotyping in Parkinson's disease: A translational missing link in disease-modifying clinical trials. Mov Disord 2019;34:319-24.
6. Ridyard CH, Hughes DA. The fragility index in randomized controlled trials as a means of optimising patient benefit. Trials 2016;17:467.
7. A Study to Evaluate Efficacy and Safety of QGE031 as Add-on Therapy. Clinical Trials.gov; (n.d.). Available from: https://clinicaltrials.gov/ct2/show/NCT02987543 [Last accessed on 2023 Jul 19].
8. Lynparza Phase III PROfound Trial in HRR Mutation-Selected Metastatic Castration-Resistant Prostate Cancer Met Primary Endpoint. United Kingdom: AstraZeneca; 2019. Available from: https://www.astrazeneca.com/media-centre/press-releases/2019/lynparza-phase-iii-profound-trial-in-hrr-mutation-selected-metastatic-castration-resistant-prostate-cancer-met-primary-endpoint-07082019.html [Last accessed on 2023 Jul 19].
9. Intercept Announces Positive Data for Fibrosis Due to NASH in New Analysis. United States: Intercept Pharmaceuticals; (n.d.). Available from: https://ir.interceptpharma.com/news-releases/news-release-details/intercept-announces-positive-data-fibrosis-due-nash-new-analysis [Last accessed on 2023 Jul 19].
10. Higgins-Dunn N. Intercept Still Sees FDA Approval Pathway for NASH Hopeful Ocaliva, as Long as Safety Holds. Fierce Pharma; (n.d.). Available from: https://www.fiercepharma.com/pharma/intercept-still-sees-fda-approval-pathway-for-nash-hopeful-ocaliva-as-long-as-safety-holds [Last accessed on 2023 Jul 19].
11. Sagonowsky E. Intercept's NASH Hopeful Turned Away by FDA, Raising Questions for Other Companies in Race. Fierce Pharma; (n.d.). Available from: https://www.fiercepharma.com/pharma/intercept-s-nash-hopeful-turned-away-by-fda-raising-questions-for-other-companies-race [Last accessed on 2023 Jul 19].
12. Estellat C, Ravaud P. Lack of head-to-head trials and fair control arms: Randomized controlled trials of biologic treatment for rheumatoid arthritis. Arch Intern Med 2012;172:237-44.
13. REVEL: The First and Only Phase III Trial to Show Superior Progression-Free Survival and Overall Survival vs Docetaxel in 2L+ Advanced or Metastatic NSCLC. Cyramza, Eli Lilly and Company; (n.d.). Available from: https://www.cyramza.com/hcp/nsclc-treatment/revel-survival-rate-efficacy [Last accessed on 2023 Jul 19].
14. Top Line Results - Lupuzor™ Pivotal Phase III Trial. ImmuPharma; 2020. Available from: https://www.immupharma.co.uk/top-line-results-lupuzor-pivotal-phase-iii-trial [Last accessed on 2023 Jul 19].
15. Gote AP. Lupuzor Shows Promising Results in Phase 3 Study in Lupus Patients. Lupus News Today; 2020. Available from: https://lupusnewstoday.com/news/lupuzor-shows-promising-results-in-phase-3-study-in-lupus-patients [Last accessed on 2023 Jul 19].
16. Feuerstein A. Sage Changes Primary Endpoint in Key Study Months After Analysts Fret about Depression Drug's Durability. Endpoints News; 2021. Available from: https://endpts.com/sage-changes-primary-endpoint-in-key-study-months-after-analysts-fret-about-depression-drugs-durability

Funding: Nil; Conflicts of Interest: None Stated.

How to cite this article: Varma N. A holistic framework for interpreting positive clinical outcomes in biotechnology. Eastern J Med Sci. 2023;8(2):30-33.