Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter now, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Lu Zhang April 15, 2021

Protocol Amendments and Probability of Success in Clinical Trials

by

Lu Zhang

Clifford J. Carrubba

Adviser

Department of Mathematics

Clifford J. Carrubba

Adviser

Yuanzhe Xi

Committee Member

Roberto Franzosi

Committee Member

2021

Protocol Amendments and Probability of Success in Clinical Trials

By

Lu Zhang

Clifford J. Carrubba

Adviser

An abstract of

a thesis submitted to the Faculty of Emory College of Arts and Sciences

of Emory University in partial fulfillment

of the requirements of the degree of

Bachelor of Science with Honors

Department of Mathematics

2021

Abstract

Protocol Amendments and Probability of Success in Clinical Trials

By Lu Zhang

During a drug's clinical trial, outcomes should be defined before the trial starts and not amended after. However, the selective publication of studies with statistically significant results and selective reporting of favorable outcomes are not uncommon. Outcome amendments are also prevalent, which can influence the trials' results. This study's primary purpose is to investigate drugs' mid-trial endpoint amendments' effect on its trials' probability of success. A large dataset containing 13584 clinical trials associated with 313 drugs is analyzed by an automated algorithm capturing and categorizing different protocol amendments. Statistical analyses demonstrate the prevalence of multiple testing and protocol amendments in clinical trials. On average, each clinical trial declared 1.91 primary outcomes and 5.93 secondary outcomes. Around 18% of protocols are amended after pre-registration. Pearson correlation test shows that protocols with one type of changes are significantly more likely to have other types of amendments. Multiple linear regression shows that the number of protocol changes, especially changes in the primary outcome, significantly affects the drug's probability of success. This study suggests that the FDA should consider enforcing better regulation of protocol changes after the start of clinical trials.

Protocol Amendments and Probability of Success in Clinical Trials

By

Lu Zhang

Clifford J. Carrubba

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences

of Emory University in partial fulfillment

of the requirements of the degree of

Bachelor of Science with Honors

Department of Mathematics

2021

Acknowledgements

I would like to give thanks to my advisor Dr. Clifford J. Carrubba. This thesis and the research would not have been possible without the exceptional support from my advisor. His knowledge, enthusiasm, and caring personality guided me through the whole research process. Dr. Matthew Gabel and Dr. Brian F. Crisp at Washington University in St. Louis inspired me when forming the research question and shared their invaluable comments with me. I am also grateful for the insightful opinions and support from Dr. Yuanzhe Xi and Dr. Roberto Franzosi on my honor committee.

I also want to thank all my friends and my parents who support me this whole time.

Table of Contents

INTRODUCTION

This honor thesis was inspired by the clinical process of Remdesivir, which was regarded as a potential cure for COVID-19 in summer 2020. Its primary endpoint changes from mortality rate to time-to-recovery in the middle of the clinical research process are particularly concerning. Such change after pre-registration should raise the flag for any statistician of p-hacking, more so in this case since the Food and Drug Administration (FDA) used the modified result as a basis to approve Remdesivir.

The primary purpose of this study is to investigate drugs' endpoint changes' effect on its trials' probability of success.

BACKGROUND

There are three phases of clinical research before the final drug approval.

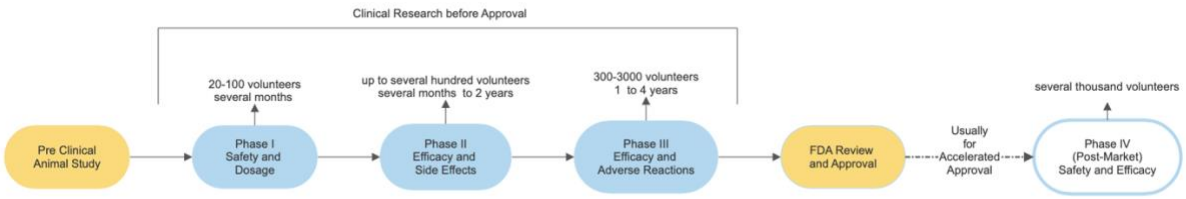


Figure 1. FDA drug approval process

According to the FDA, Phase I study is the first time researchers test the safety and safe dosage range of the experimental drug in a small group of volunteers, typically 20 to 100. This phase usually lasts for several months. Once it passes through Phase I, the drug is given to a larger group of volunteers, which can be up to several hundred, in Phase II. This phase examines

the efficacy of the drug and further evaluates its safety and side effects. It usually lasts for several months to 2 years. If the test drug can move to the next phase, it will be administered to an even larger group of volunteers, ranging from 300 to 3000, in Phase III. Researchers must confirm the drug's efficacy, monitor its adverse reactions, and compare it to the commonly used treatments in this phase. The length of study is usually 1 to 4 years.

Controlled clinical research can have both primary and secondary outcomes. Clinical outcomes, or clinical endpoints, are objective outcome measures used to determine whether the studied intervention shows substantial evidence of benefits (National Cancer Institute). An example of clinical outcomes is the overall survival based on the death from any cause. Primary outcomes are the most relevant variables in answering the research question. Secondary outcomes are selected to demonstrate additional effects of the drugs (FDA). All clinical outcomes should be defined before the trial starts. The results are likely biased if the outcomes are decided or amended after.

LITERATURE REVIEW

Relevant previous research mainly focused on two areas. One is the selective publication of studies with statistically significant results. That is, trials with statistically significant findings are more likely to be published, and often more quickly, than trials with non-significant ones (Hopewell et al. 2009).

The other area is the selective reporting of favorable outcomes within studies. According to Chan et al.'s manual examination of 122 published journal articles and their corresponding clinical protocols, more than half of outcomes per trial were reported incompletely; 62% of trials had at least one primary outcome amended. A later systematic review showed significant

heterogeneity in the discrepancy rate between the registered outcomes and published outcomes, ranging from less than 10% to more than 60% (Jones et al. 2015). The cause of such heterogeneity is unknown. Previous studies have limitations: First, due to the need for close reading, the sample size is usually limited to around two hundred. Researchers can thus only focus on a specific disease area. This sample size and data selection can reveal a general trend but may not be enough to generalize to all clinical trials. Second, some scholars selected journal articles first and searched for their corresponding clinical protocols, ignoring unpublished trials altogether. The selective publication process may lead to biases in the final results. Last, the manual comparisons between published journal articles and clinical protocols are highly subjective. Methodological differences can affect outcome consistency.

Outcome changes within the clinical research process, or protocol amendments, were further studied by a small number of scholars. However, those studies emphasize a basic description of the average occurrence rate (Getz et al. 2011) and potential economic impact (Getz et al. 2016). To the best of this author's knowledge, no previous research has empirically studied the impact of protocol amendments on drugs' approval rate.

This study analyzes two inter-related databases containing more than 350,000 trials, a large sample size. An objective standard to categorize different types of protocol amendments was created. An automated algorithm to capture protocol amendments and link them with drugs' probability of approval was developed.

DATA

Two datasets are used in this study: ClinicalTrials.gov and Citeline.

To encourage transparency and consistency in reporting the clinical trial outcomes, ClincialTrial.gov was first made available to the public in 2000. In 2005 the International Committee of Medical Journal Editors (ICMJE) required all publications to have pre-registered trial protocols. In the same year, the State of Maine passed a law requiring prescription drug manufacturers or labelers to submit both clinical registration and results to ClinicalTrials.gov. The FDA Amendments Act of 2007 requires all clinical trials to be registered and amendments to be tracked via Clinical trials.gov. As a result of the policies, trial registration has dramatically increased, and the number of missing data elements has generally declined (Califf et al. 2012). To assess the consistency between different versions of protocols in individual clinical trials, data used in this study are scraped from ClinicalTrials.gov Archive containing individual studies' history of changes.

Access to the Citeline database was gained with support from China Pharmaceutical University. Provided by Informa Pharma Intelligence, this database collates commonly used clinical data sources, including ClinicalTrials.gov, study reports, institutional press releases, and drug marketing label applications. It contains trial information both in and outside the United States. Citeline data is a superset of the Trialtrove database and the Pharmaprojects database. Trialtrove provides information about individual clinical trials, which can be matched with ClinicalTrials.gov data by a unique identifier called NCT number. Pharmaprojects integrates all drug approval data. Pharmaprojects is linked with Trialtrove using an identification code for all Citeline Drug ID. The Citeline data used includes one unique observation per drug per indication (drug-indication pair). As a result, a single trial that tests multiple outcomes can be repeated to

form multiple observations. It is not uncommon for the database to contain missing data points. This dataset was used to measure drugs' probability of success.

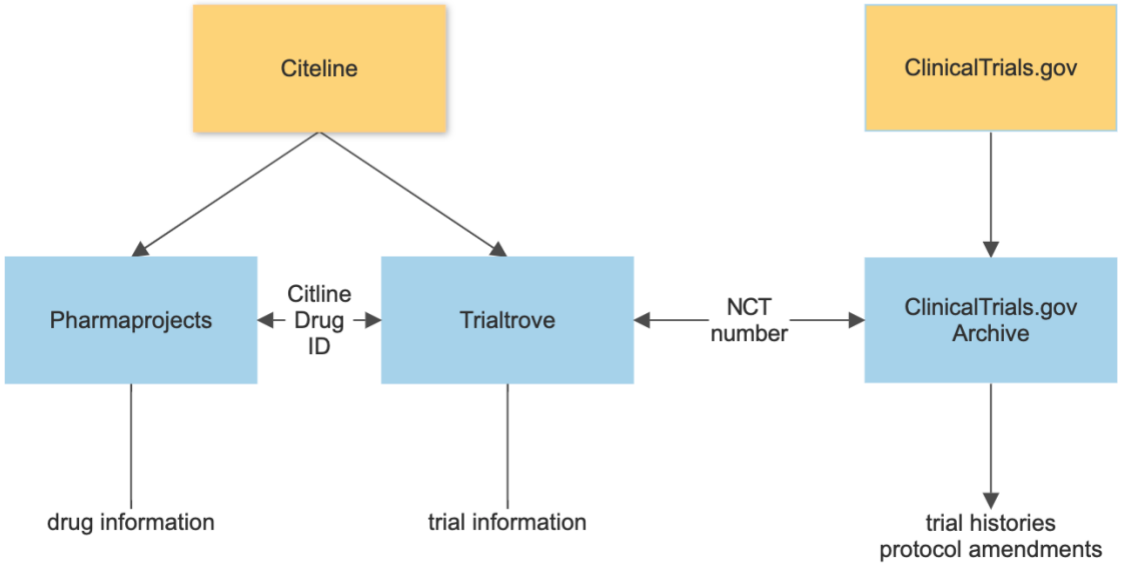


Figure 2. Summary of the databases

Due to limited access to the Citeline database, only drugs related to bladder cancer were included in this study. A total of 313 drugs are selected, with 24742 unique trials and 46614 unique observations. As stated above, ClinicalTrials.gov is a subset of the Citeline data. Consequently, only 14394 in the 24742 trials (58.18%) are registered on Clinicaltrials.gov. The historical versions of 810 (5.63%) registered trials are not recorded in the Archive. In total, 13584 clinical trials associated with 313 drugs are included in the final dataset. It is not a random sample: The generalization of results should be insightful, but the external validity is not guaranteed. However, since the automated algorithm is applied in this study, a more comprehensive analysis of overall recorded clinical trials can be done with complete access to datasets.

METHODOLOGY

This study's primary goal is to test if drugs' clinical trial outcomes were selected or modified after pre-registration to achieve a higher probability of success. Two main assessed factors are protocol amendments and probability of success.

Ten types of protocol amendments are evaluated: 1) Outcome changes from primary to secondary; 2) Outcome changes from secondary to primary; 3) Primary outcome omitted from the protocol; 4) Secondary outcome omitted from the protocol; 5) Primary outcome amended; 6) Secondary outcome amended; 7) New primary outcome introduced; 8) New secondary outcome introduced; 9) Timeframe of what the primary outcome is assessing amended (e.g., changing from "overall survival rate in 2 years" to "overall survival rate in 12 months"); 10) Timeframe of what the secondary outcome is assessing amended. An automated algorithm was developed to scrape and compute the total number of amendments per clinical trial by type.

The overall Probability of Success (POS) (Wong, Siah, and Lo 2019) moving a drug from Phase 1 to approval (POS1, APP) is the same as the likelihood of approval (LOA) (Hay et al. 2014). Previous research, regardless of the scope, focuses more on strategic decisions. That is to say, the primary focus of the research is to provide up-to-date information to potential investors. This POS reduces the risk of investors by allowing them to make more scientific and economic decisions. Due to the differences in years covered, methodology, and sample source, the results vary among publications. However, there are general trends merged. The general POS from Phase I to final approval ranges from 9.6% to 13.8%. (Thomas et al. 2016, Wong, Siah, and Lo 2019, Smietana et al. 2016, Hay et al. 2014)

The previous predominant method is "phase-by-phase." This method computes POS by first calculating the ratio of observed phase transitions to observed drug development programs

in that phase; then, multiply this ratio with individual phase probabilities. However, missing data points in Phase II would decrease the number of phase transitions and decrease the overall ratio. This phase-by-phase method tends to underestimate the POS. Another method known as "path-by-path" was introduced and used by Smietana et al. (2016) and Wong, Siah, and Lo (2019) to cope with the missing data. If a drug-indication pair directly advanced from Phase I to Phase III, this method assumes that at least one Phase II trial occurred but is missing from the dataset. It takes the in-progress trials into account as well. Wong's team simulated POS using the path-by-path approach, and it accurately estimated the POS with missing phase transitions.

Given the advantages of the path-by-path approach, an algorithm similar to the algorithm in Wong, Siah, and Lo (2019)'s study is constructed. Standard assumptions were made: Phase I/II and Phase II/III trials are considered Phase II and Phase III, respectively. If multiple clinical trials occurred for a single drug-indication pair, the number of changes is added together by types.

After examining the dataset's basic statistics, a multiple linear regression was performed to search for significant factors impacting the POS of a particular phase for the drug-indication pair.

RESULTS

Prevalence of Multiple Outcomes

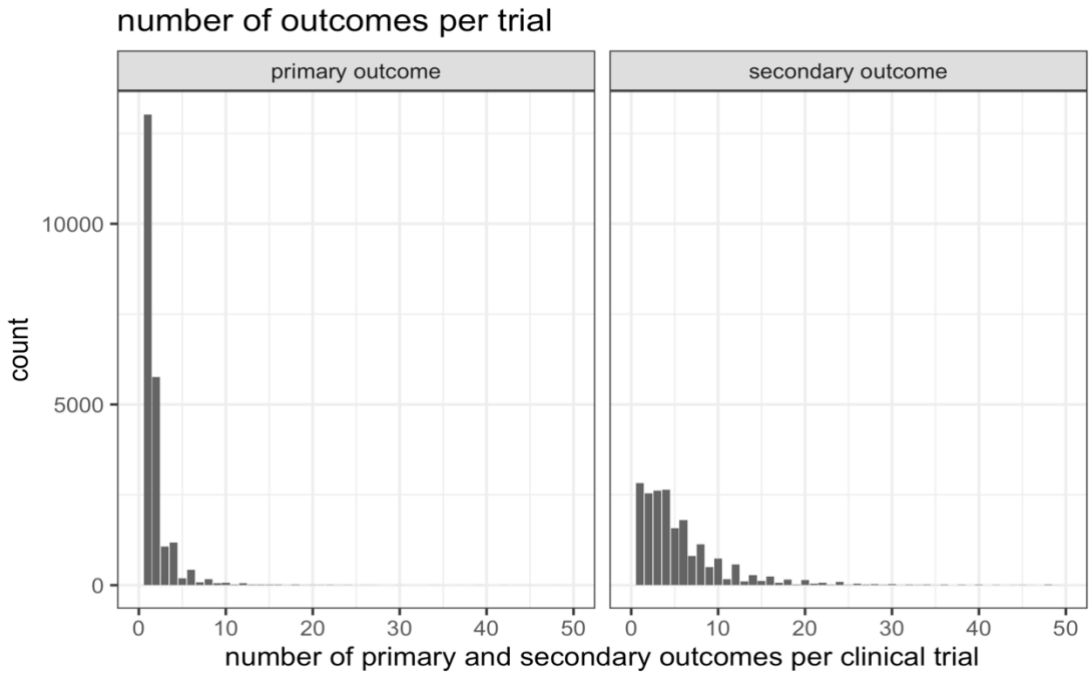
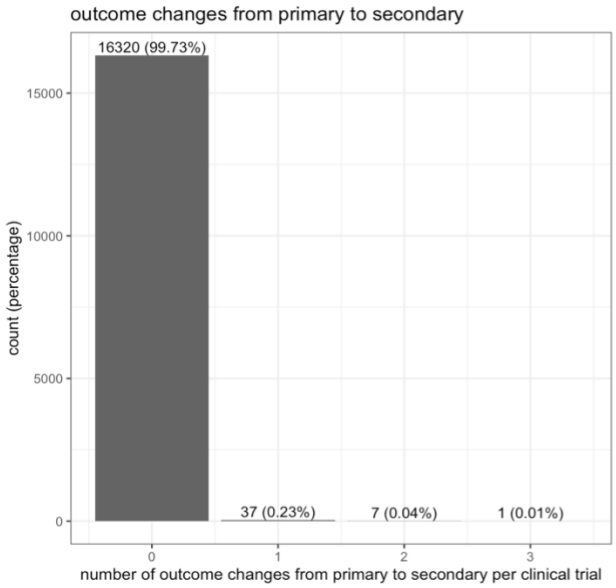
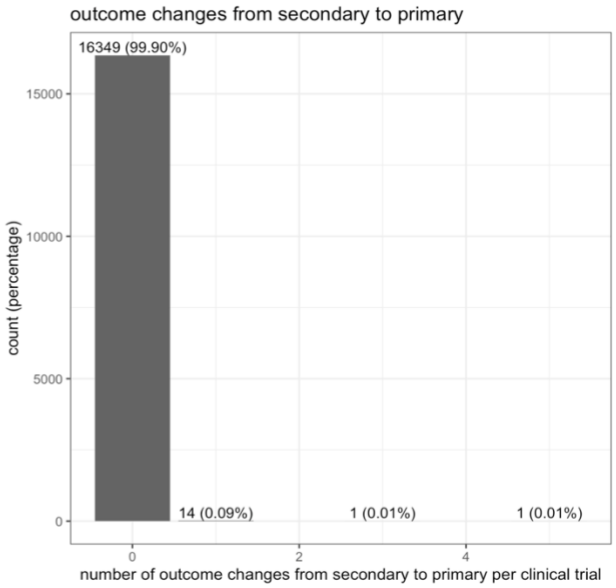


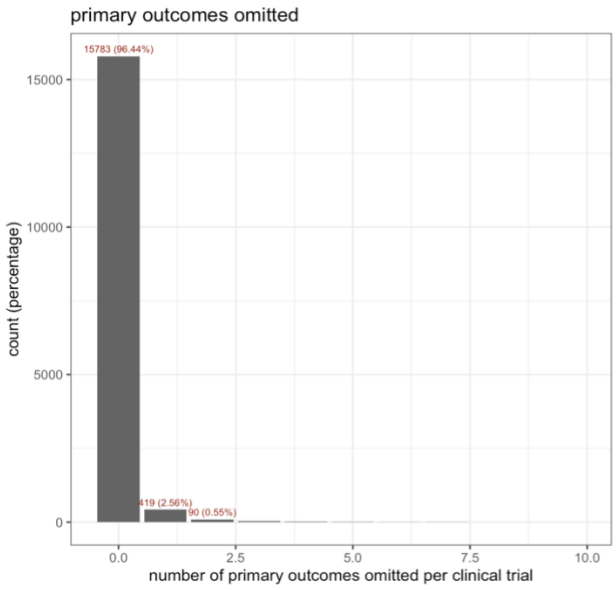
Figure 3. Number of proposed outcomes per clinical trial.

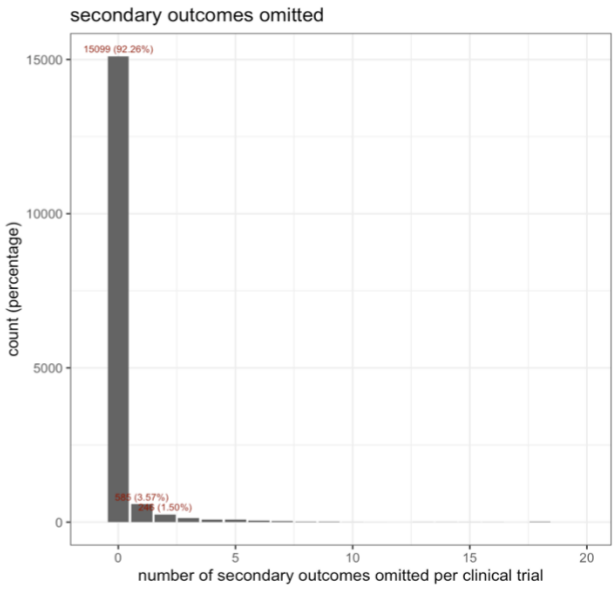
On average, the surveyed clinical trials have 1.91 primary outcomes and 5.93 secondary outcomes. Because the success of any one of the primary outcomes can prove the efficacy of the test drug, the rate of falsely concluding the drug's efficacy increases due to multiple outcomes. All publications with multiple outcomes should perform statistic correction. However, very few studies adjusted for multiple outcomes (Tyler, Normand, and Horton 2011, Vickerstaff et al. 2015), and it is not currently listed as a condition for publication by the FDA.

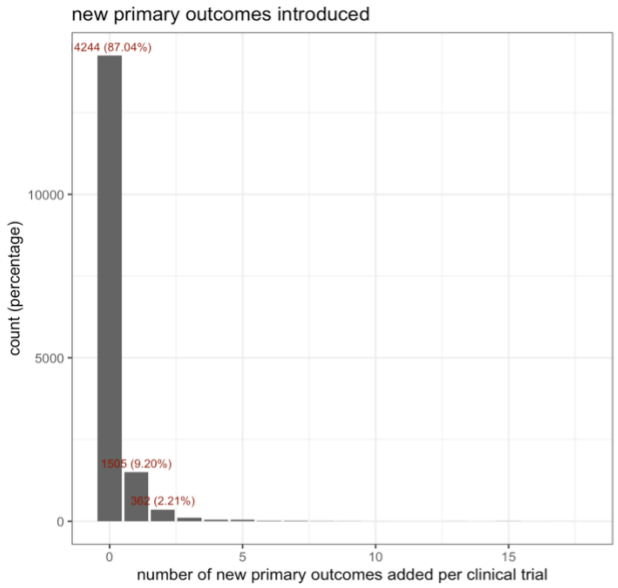
Prevalence of Protocol Amendments

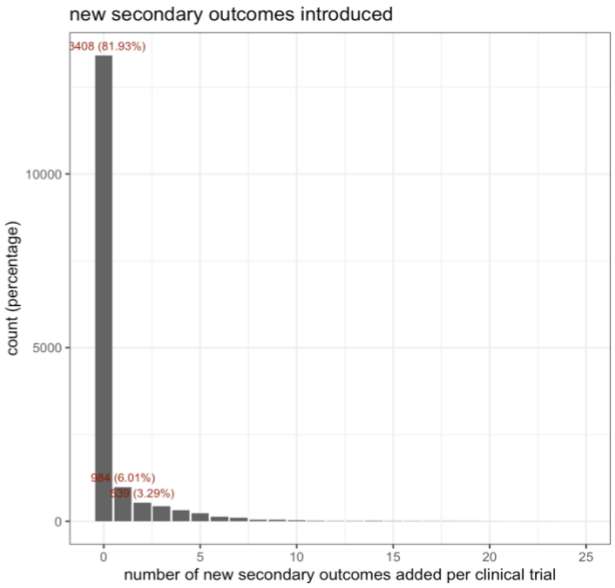


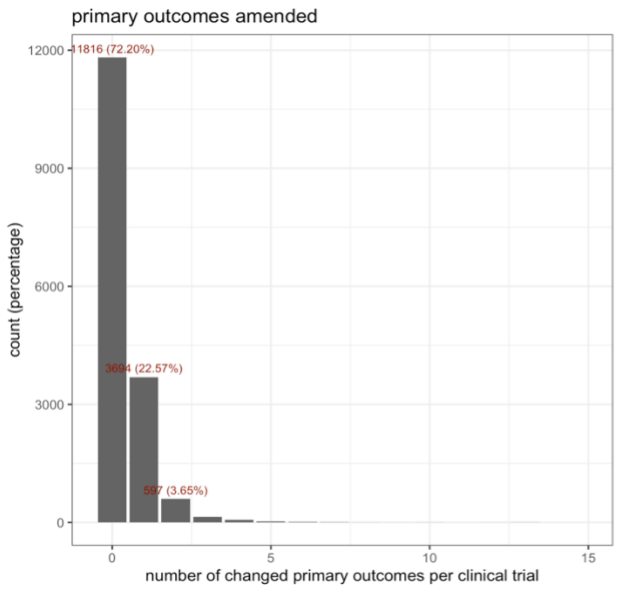


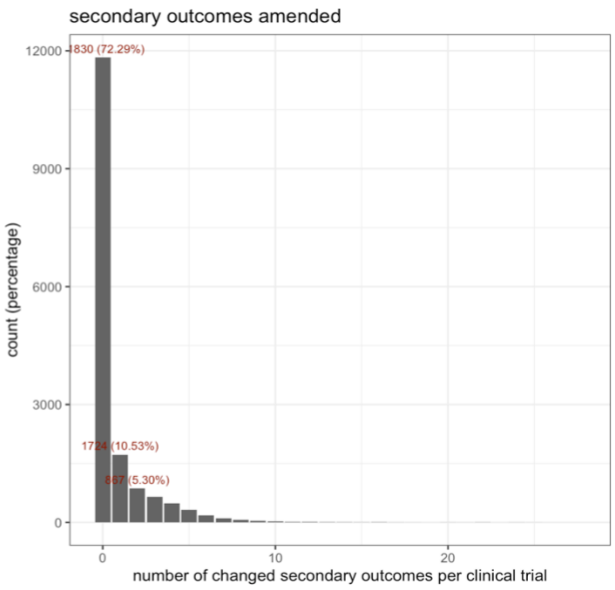





new primary outcomes introduced
4244 (87.04%)
1505 (9.20%)
362 (2.21%)
count (percentage)
number of new primary outcomes added per clinical trial


new secondary outcomes introduced
3408 (81.93%)
934 (6.01%)
538 (3.29%)
count (percentage)
number of new secondary outcomes added per clinical trial


primary outcomes amended
11816 (72.20%)
3694 (22.57%)
597 (3.65%)
count (percentage)
number of changed primary outcomes per clinical trial


secondary outcomes amended
11830 (72.29%)
1734 (10.53%)
867 (5.30%)
count (percentage)
number of changed secondary outcomes per clinical trial

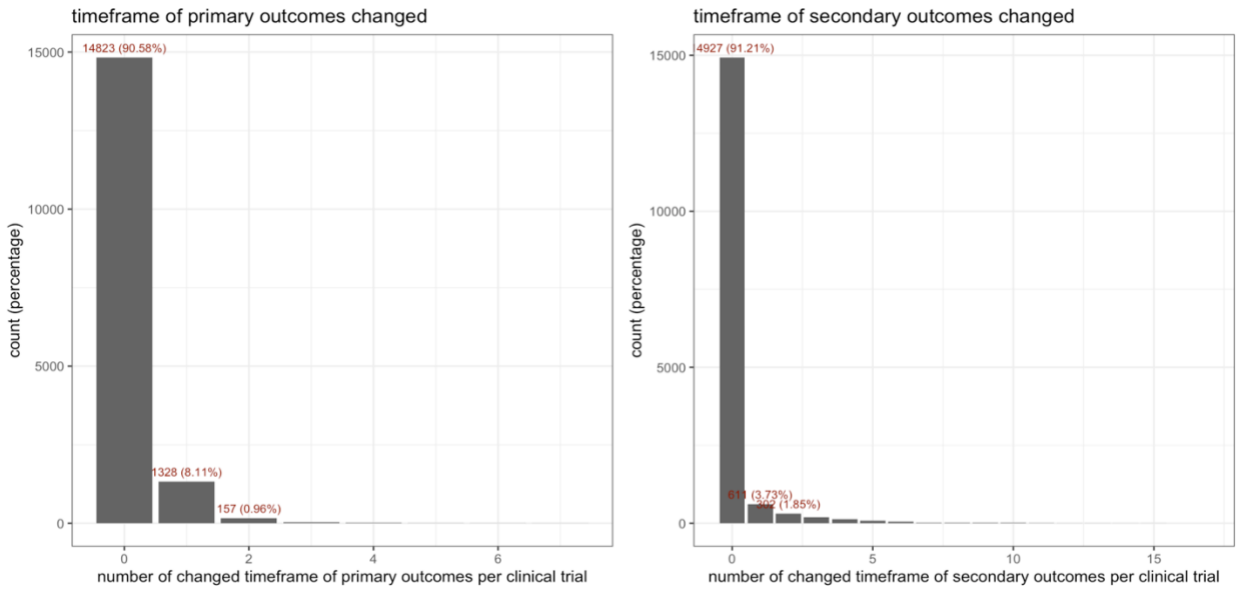
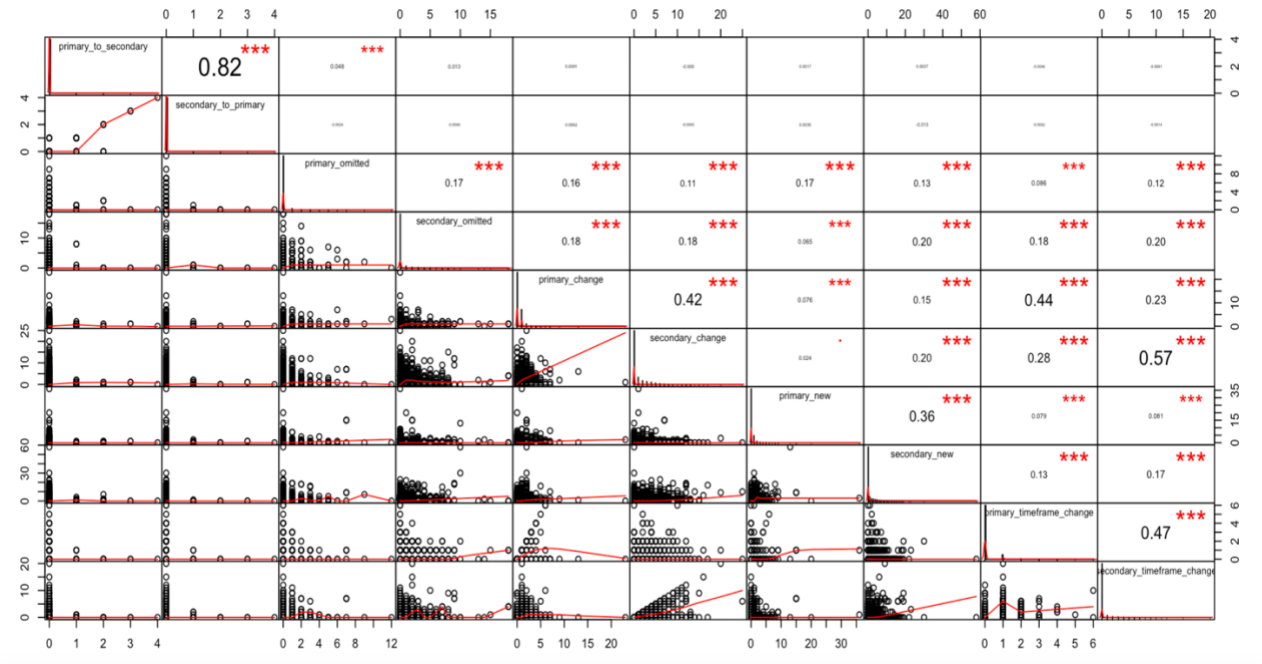


Figure 4. Number of protocol amendments by type.

Outliers were excluded in the visualization to reveal the general distribution, but those outliers were included in all analyses performed. According to the data distribution in figure 4, there is a prevalence of protocol changes during the clinical research process: 12.96% of clinical trials have new primary outcomes introduced, and around 18.07% have secondary outcomes introduced. Primary and secondary outcomes are amended in 27.80% and 27.71% of the study, respectively. Their timeframes are changed in 9.42% and 8.79% of the study, respectively.

Correlation Matrix for different types of protocol amendments



***p<.001. **p<.01. *p<.05.

Figure 5. Correlation matrix for different types of protocol amendments

A Pearson correlation test is performed to analyze the correlation between different types of protocol amendments. In Figure 5, numbers in the upper-triangle are the Pearson correlation coefficients. Graphs in the lower-triangle visualize the linear relationships between each pair of variables. Variable names are listed in the diagonal.

Except for outcome changes from primary to secondary and from secondary to primary, other protocol amendments are statistically significantly correlated. All the corresponding correlation coefficients are positive: protocols with one type of amendments are more likely to have other types of amendments. Outcome changes from primary to secondary and from secondary to primary are strongly correlated with each other. Researchers usually swap primary and secondary outcomes. Moreover, changing primary and secondary outcomes, introducing new

primary and secondary outcomes, and changing timeframes of primary and secondary outcomes are moderately correlated with each other. This indicates that researchers tend to apply the same protocol amending method to both primary and secondary outcomes. Both primary and secondary outcome changes are strongly correlated to their timeframe changes: When modifying outcomes, researchers also adjust related timeframes.

Probability of Success

Table I. Probability of Success by Phases

	POS_12 (%) probability of success moving from phase 1 to phase 2	POS_23 (%) probability of success moving from phase 2 to phase 3	POS_3a (%) probability of success moving from phase 3 to final approval	POS_1a (%) overall probability of success moving from phase 1 to final approval
Bladder Cancer	78.2	23.6	49.3	9.1

The overall probability of success (POS_1a) lies within the general POS range listed in the methodology section. Most trials fail when moving from Phase II to phase III.

Multiple Linear Regression

Multiple linear regression was performed on the POS of phase transitions and the overall POS for every drug-indication pair. For the first three models, the dependent variable is the POS moving from phase i to i+1, and the independent variables are the number of different types of protocol amendments during phase i. The regression equation is:

$$POS_{i,i+1} = \beta_1\, primary\ to\ secondary + \beta_2\, secondary\ to\ primary + \beta_3\, primary\ amended + \beta_4\, secondary\ amended + \beta_5\, new\ primary\ introduced + \beta_6\, new\ secondary\ introduced + \beta_7\, primary\ omitted + \beta_8\, secondary\ omitted$$

The model for overall POS moving from phase I to final approval is similar. The number of protocol amendments is calculated by aggregating all protocol amendments in previous phases.

Timeframe changes are not included in the regression model since how it affects the clinical outcomes can differ from amending the text.

Table II. Multiple Linear Regression Result

	Model on POS_12			Model on POS_23			Model on POS_3a			Model on POS_1a		
Variable	Estimate	SE	p-value	Estimate	SE	p-value	Estimate	SE	p-value	Estimate	SE	p-value
primary to secondary	-2.74	7.61	0.719	1.27	3.61	0.724	-9.31	10.23	0.363	0.92	2.41	0.701
secondary to primary	-1.86	11.30	0.870	-9.27	11.36	0.414	10.25	21.52	0.634	-1.79	4.17	0.668
primary amended	0.99	0.30	0.001***	-0.98	0.34	0.817	-1.99	1.00	0.047*	-0.60	0.17	0.0004***
secondary amended	0.48	0.17	0.005**	0.30	0.13	0.030*	0.43	0.31	0.165	0.11	0.07	0.120
new primary introduced	0.88	0.21	2.2e-05***	-0.97	0.31	0.002**	-0.97	0.82	0.180	-0.43	0.13	0.001**
new secondary introduced	-0.14	0.11	0.196	-0.12	0.14	0.393	-1.29	0.22	0.435	0.02	0.06	0.791
primary omitted	-1.72	0.61	0.005**	-2.10	0.84	0.012*	0.93	0.47	0.378	-1.60	0.37	1.62e-05***
secondary omitted	0.99	0.13	1.6e-13***	0.40	0.23	0.085	-0.010	0.60	0.122	0.79	0.09	<2e-16***
R^2		0.075			0.011			0.050			0.020	
F-statistic		27.9***			3.854***			4***			17.39***	

****p*<.001. ***p*<.01. **p*<.05.

The number of protocol amendments is significantly related to the POS in earlier phases and the overall POS. However, these changes' direction of effects is affected by various compounding factors, including researchers' motives when changing their proposed outcomes. Researchers may amend their protocols to get a lower p-value before the clinical trial's

completion, or they may fish for an outcome success by wildly changing their protocols on a clinical trial that they deemed destine to fail. Swapping primary and secondary outcomes are generally unrelated to the POS. This may be a result of the low occurrence rate of such events. The low F-statistic but high R suggest that although the model is a good fit for the dataset, the current model cannot describe many dataset variances. Other potential explanatory factors not considered in this study include the duration of clinical trials, number of tested primary and secondary outcomes, and number of clinical trials in a given drug-indication pair.

CONCLUSION AND DISCUSSION

Multiple outcome testing and protocol amendments are both prevalent in clinical trials. On average, trials have around two primary outcomes and six secondary outcomes, while only a few published results perform any statistical correction. Approximately 18% of surveyed clinical trials have amended primary and secondary outcomes. Around 10% of the clinical trials change the timeframe of primary and secondary endpoints.

Any protocol amendments after pre-registration can create bias to the final results. The Pearson correlation matrix result shows that except for outcome changes from primary to secondary and from secondary to primary, other protocol amendments are positively correlated. Protocols with one type of amendments are more likely to have other types of amendments. Clinical researchers usually change primary outcome to secondary and change secondary outcome to primary at the same time. They also tend to apply the same protocol amending method to both primary and secondary outcomes. They usually choose to introduce a new outcome, omit an old outcome, or amend existing outcomes instead of using all three protocol amending methods. When modifying the outcomes, researchers also adjust related timeframes.

The overall POS from Phase I to final approval of drugs for bladder cancer is 9.1%. Most of the drugs fail in Phase II. Number of protocol changes significantly affects the POS of the clinical trial. Primary outcomes are more closely related to the POS, since success with only secondary outcomes cannot move clinical research to the next phase. The intention of protocol amendments is still unclear. The proposed models are not predictive, and other explanatory variables can be added to describe the POS better. The protective mechanism of pre-registration required by the FDA is not as effective if a considerable number of protocol changes occur, especially since those changes are significantly related to the trials' success rate. Better regulation on the report of protocol changes should be enforced.

More investigations can be done on this dataset. For instance, the difference-in-difference method can be implemented to test if the 2007 FDA requirement of trial pre-registration alters protocol changing patterns and the POS. Timeframe changes can be added to the model, as well as sponsor type. The false-positive rate can be calculated based on the number of outcomes per clinical trial. Similarity scores of outcomes before and after amendments have been recorded by the algorithms, and analysis on how the percentage of outcome changes affect the POS can yield useful information.

REFERENCES

Califf, Robert M., Deborah A. Zarin, Judith M. Kramer, Rachel E. Sherman, Laura H. Aberle, and Asba Tasneem. 2012. "Characteristics of Clinical Trials Registered in ClinicalTrials.Gov, 2007-2010." *JAMA* 307(17):1838–47.

Getz, Kenneth A., Stella Stergiopoulos, Mary Short, Leon Surgeon, Randy Krauss, Sybrand Pretorius, Julian Desmond, and Derek Dunn. 2016. "The Impact of Protocol Amendments on Clinical Trial Performance and Cost." *Therapeutic Innovation & Regulatory Science* 50(4):436–41.

Getz, Kenneth A., Rachael Zuckerman, Anne B. Cropp, Anna L. Hindle, Randy Krauss, and Kenneth I. Kaitin. 2011. "Measuring the Incidence, Causes, and Repercussions of Protocol Amendments." *Drug Information Journal : DIJ / Drug Information Association* 45(3):265–75.

Hay, Michael, David W. Thomas, John L. Craighead, Celia Economides, and Jesse Rosenthal. 2014. "Clinical Development Success Rates for Investigational Drugs." *Nature Biotechnology* 32(1):40–51.

Hopewell, Sally, Kirsty Loudon, Mike J. Clarke, Andrew D. Oxman, and Kay Dickersin. 2009. "Publication Bias in Clinical Trials Due to Statistical Significance or Direction of Trial Results." *The Cochrane Database of Systematic Reviews* (1):MR000006.

Jones, Christopher W., Lukas G. Keil, Wesley C. Holland, Melissa C. Caughey, and Timothy F. Platts-Mills. 2015. "Comparison of Registered and Published Outcomes in Randomized Controlled Trials: A Systematic Review." *BMC Medicine* 13(1):282.

National Cancer Institute. 2011. "Definition of Endpoint - NCI Dictionary of Cancer Terms - National Cancer Institute." Retrieved April 8, 2021

Smietana, Katarzyna, Marcin Siatkowski, and Martin Møller. 2016. “Trends in Clinical Success Rates.” *Nature Reviews. Drug Discovery* 15(6):379–80.

Tyler, Kristin M., Sharon-Lise T. Normand, and Nicholas J. Horton. 2011. “The Use and Abuse of Multiple Outcomes in Randomized Controlled Depression Trials.” *Contemporary clinical trials* 32(2): 299–304.

Vickerstaff, Victoria et al. 2015. “Are Multiple Primary Outcomes Analysed Appropriately in Randomised Controlled Trials? A Review.” *Contemporary Clinical Trials* 45: 8–12.

Wong, Chi Heem, Kien Wei Siah, and Andrew W. Lo. 2019. “Estimation of Clinical Trial Success Rates and Related Parameters.” *Biostatistics* 20(2):273–86.