

Effects of biomarker diagnostic accuracy on biomarker-guided phase 2 trials

Jay JH. Park^{a,b,*}, Ofir Harari^b, Louis Dron^b, Edward J. Mills^{b,c}, Kristian Thorlund^{b,c}^a Experimental Medicine, Department of Medicine, University of British Columbia, Vancouver, BC, Canada^b MTEK Sciences, Vancouver, BC, Canada^c Department of Health Research Methodology, Evidence, and Impact, McMaster University, Hamilton, ON, Canada

ARTICLE INFO

Keywords:

Biomarker-guided trials
Phase II trials
Simon's two-stage design
Biomarker test accuracy

ABSTRACT

Recent advancements in genomics have attracted attention towards biomarker-guided trials. These trials aim to identify therapies that target diseases based on their genetic profile, and are especially common in cancer research. Careful incorporation of biomarkers in phase II studies is critical to the selection of candidates for further phase III investigation. This short communication focuses on problems of biomarker test accuracy in biomarker-guided trials. We assessed how diagnostic accuracy of biomarker tests affects type I error rate, statistical power, and sample size requirements of single-arm biomarker-guided trials. In particular, we report how false positive rates (FPRs) of biomarker tests reduce statistical power and type I error for Simon's two-stage design, and the degree of sample size correction required to achieve pre-specified power and type I error with varying FPRs. This was done using a case study based on a previous biomarker-guided single-arm trial that was designed with an assumed tumor response rate of 10% under the null hypothesis and 40% for the alternative hypothesis for the mutant group for 5% type I error and 90% power. With varying FPRs of biomarker tests, we considered two scenarios in which the response rate for the wild-type group was assumed to be lower than the response rate for the mutant group at 5% and 10%. We also developed a simple open-source online trial planner for future investigators to use for their biomarker-guided phase II trials (https://mtek.shinyapps.io/Biomarker_Trial_Planner/).

1. Background

Recent advancements in genomics have improved our understanding of the genetic landscape of different diseases, especially within oncology [1]. Biomarker-guided trials for precision medicine, which aim to identify targeted therapies for specific genetic profiles, have emerged as an area of increased interest [2,3]. While there is much promise in this targeted approach in terms of our ability to treat diseases, investigators should be aware of the several pitfalls to this technique. In particular, investigators need to consider the inaccuracy that accompanies all medical tests and therefore the proportion of incorrectly-classified bio-marker negative patients that can be anticipated in their trials. Previous reviews of biomarker-guided trial designs [4,5] and guidance documents by the Food and Drug Administration (FDA) [6,7] have all co-highlighted the importance of biomarker test accuracy in biomarker-guided clinical trials. While many published trials have mentioned that the measures undertaken to confirm biomarker test accuracy, translating this accuracy in context for the operational characteristics of these trials has been limited.

Careful incorporation of biomarkers in phase II studies is critical to the selection of candidates for further phase III investigation. In phase II

oncology trials, the general goal is to assess whether a treatment has sufficient activity against a specified tumor type and therefore whether further investigation is warranted. Thus, these exploratory trials should be designed to maximize statistical power, while confirmatory studies should be designed to minimize type I errors [8]. As patients are recruited on the basis of biomarker test in biomarker-guided trials, the accuracy of the test itself in combination with the influence it has on treatment effect are potentially pivotal in determining whether a study and treatment are to be considered successful.

The purpose of this short communication is two-fold. Firstly, we assessed the effects of biomarker testing accuracy on *type I error rate, power, and sample size* of single-arm phase II clinical trials with Simon's two-stage design. Simon's two-stage design was chosen due to its common employment in phase II oncology trials [9]. We present on a single case study based on a previously conducted biomarker-guided single-arm trial that used Simon's two-stage minimax design [10]. For this, we considered varying false positive rates (FPRs) of biomarker tests in two scenarios in which the response rate for the wild-type (false positive) group was assumed to be lower than the rate in the mutant (true positive) group. Secondly, we introduce a simple open-source online trial planner so that future investigators can use to plan their

* Corresponding author. Vancouver General Hospital (VGH) Research, Pavilion 575West 10th Avenue, Vancouver, BC, V5Z 1L8, Canada.

E-mail address: jayhpark1@alumni.ubc.ca (J.J. Park).

<https://doi.org/10.1016/j.conctc.2019.100396>

Received 11 April 2019; Received in revised form 7 June 2019; Accepted 13 June 2019

Available online 14 June 2019

2451-8654/ © 2019 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

exploratory biomarker-guided trials.

2. Methods

2.1. Case study general assumptions

We considered varying FPRs of biomarker tests from completely accurate to inaccurate (FPR = 0.00 to 1.00) to estimate how biomarker testing accuracy affects the type I error rate, power, and sample size for Simon's two-stage design. We assumed that tumor response rate for an experimental therapy would be higher for those with the targeted mutation (mutant group) versus those without (wild-type group). Response rates of 10% under the null hypothesis and 40% under the alternative hypothesis were assumed for the mutant group. These effect sizes were based on a previously conducted biomarker-guided phase II trial [10]. For the wild-type group, response rates of 5% and 10% were assumed as separate scenarios. Further details on hypothesis testing is provided in the Appendix.

2.2. Simon's two-stage design

In Simon's two-stage design, a trial is conducted in two stages with the option to stop the trial after the first or the second stage depending on the response observed [11]. Simon's two-stage design assumes an overall maximum number of patients (n), with a subset recruited for the first stage (n₁). The trial is stopped early for futility if the minimum pre-specified tumor response is not observed at the first stage (r₁). If the minimum response is observed, the trial continues recruitment up to n; at the second stage, the null hypothesis is rejected if the overall number of responders exceeds r. Simon's two-stage design may be designed to minimize the maximum sample size ('minimax design') or the expected sample size under the null hypothesis ('optimal design') [11].

In this short communication, we present the results on the minimax design as this design was used in the case study of Li et al., 2018 [10]. The details of our statistical analyses and the optimal design results are provided in the Appendix. Using the general assumptions on expected response rate for the target (true positive) and non-target (false positive) mutation group as mentioned before, we arrived at a minimax two-stage design with the following specifications of (n, n₁, r, r₁) = (18, 12, 4, 1), with 5% type I error rate and 90% power assuming FPR of 0.00.

3. Results

3.1. Power and type I error

Fig. 1 illustrates the effects of biomarker test's FPR on statistical power (left panel) and type I error rate (right panel) of the Simon's two-stage minimax design. The effects of the biomarker test's inaccuracy on statistical power were more prominent when a lower response rate (5%) was assumed for the wild-type. With an FPR of 10%, power was reduced from 90% to 84% when a 5% response rate was assumed for the wild-type and to 85% when a 10% response rate was assumed (Supplementary Table 1). The effects of biomarker test inaccuracy became more prominent when the FPR was greater than 10%. For instance, when the FPR was at 25%, the power became as low as 70%, with the 5% tumor response rate assumed for the wild-type group.

In addition to statistical power becoming lower than what the trial was designed with the increasing FPR, the effective type I error also became lower. When the biomarker test was completely accurate, the observed type I error was 2.77% but in all other scenarios the observed type I error was even lower (Fig. 1).

3.2. Sample size correction

Sample size correction required for the Simon's two-stage minimax design is shown in Fig. 2. When the biomarker test was completely accurate (FPR = 0.00), the maximum sample size required for 5% type I error rate and 90% power under the null hypothesis was [18]. However, the requirements for sample size increased with increasing FPR. With a fairly accurate biomarker test (10% FPR), the calculated maximum sample size was 20 for both response rates (5% and 10%) for the wild-type (Supplementary Table 2). When the biomarker test's inaccuracy was larger than 10%, the sample size correction that was required became considerable. For instance, with 27.5% FPR, the required sample size increased by over three-quarters (n = 33) with 5% response rate assumed for the wild-type.

4. Discussion

To our knowledge, this is the first study to report the effects of biomarker testing on Simon's two-stage design phase II trials. We demonstrate that increasing FPR can drastically reduce the power and type I error of a given phase II design. Testing accuracy should be therefore incorporated into trial planning. When the diagnostic test accuracy is not accounted for in the trial design, biomarker-guided

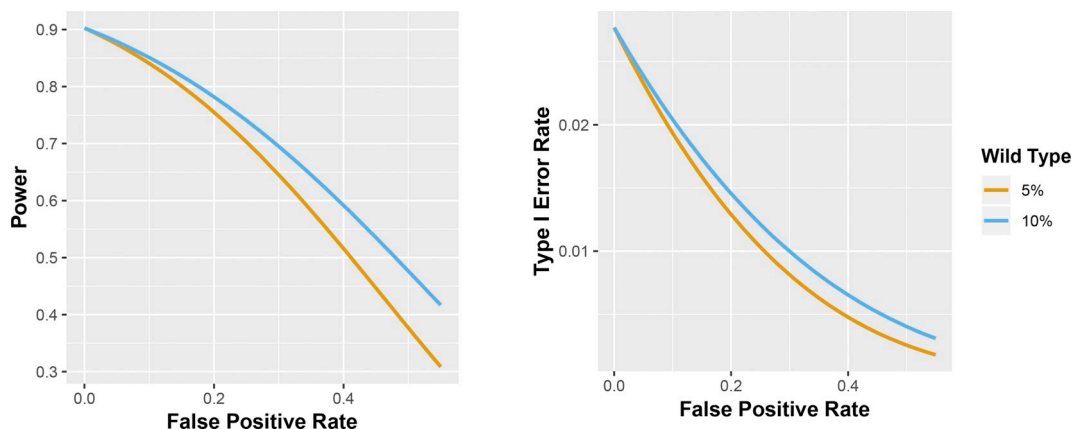


Fig. 1. Statistical power and type I error rate vs FPR, Simon's two-stage minimax design. The left panel illustrates statistical power vs false positive rates (FPR) for the Simon's two-stage minimax design [11]. The right panel illustrates type I error of the trial vs FPR of the biomarker test for the same design. We assumed the tumor response rates of 10% under the null hypothesis and 40% under the alternative hypothesis for the mutant group. For the wild-type (non-mutant) group, we assumed lower response rates of 5% and 10% (see the headers, "Wild Type"). The maximum sample size of [18] is required for 5% type I error rate and 90% power when the biomarker test is completely accurate (FPR = 0.00).

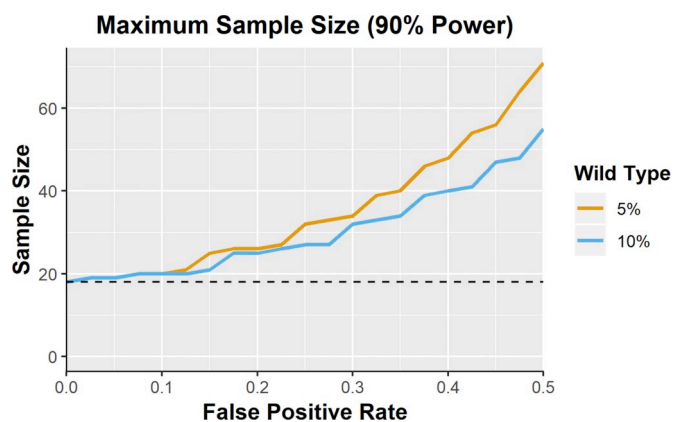


Fig. 2. Sample size required for 90% power vs FPR, Simon's two-stage minimax design

This figure illustrates sample size required for 90% power of the trial vs false positive rate (FPR) of the biomarker test for the Simon's two-stage minimax design that aims to minimize the maximum sample size [11]. We assumed the tumor response rates of 10% under the null hypothesis and 40% under the alternative hypothesis for the mutant group. For the wild-type (non-mutant) group, we assumed lower response rates of 5% and 10% (see the headers, "Wild Type"). The maximum sample size of 18 is required for 5% type I error rate and 90% power when the biomarker test is completely accurate (FPR = 0.00). The sample size required for 90% power when the biomarker test is not completely accurate (FPR > 0.00) is shown in the Y-axis.

trials will become more 'conservative', and as a result, 'effective' interventions may erroneously be deemed ineffective. As phase II studies should be designed to maximize power given their exploratory nature [8], not accounting for test accuracy in the design of these trials is particularly problematic.

To catalyze the incorporation of biomarkers into clinical trial research, we have developed a simple open-source online trial planner for future investigators to use to plan their biomarker-guided phase II trials (https://mtek.shinyapps.io/Biomarker_Trial_Planner/). This planner allows users to adjust sample sizes based on biomarker test accuracy for Simon's two-stage design for both minimax and optimal design as well as single-stage design where one final analysis is conducted at the end. We also have made our R codes publicly available for others to use and improve upon.

We recognize that there are limitations to our study. Our case study was limited to one simple case scenario. The influence of test accuracy is significantly varied depending on assumptions with regards to treatment effects, response rates in biomarker negative participants and test accuracy. To minimize this lack of generalizability, we recommend use of our open-source trial planner to tailor the findings of our work to specific disease topics and scenarios.

With the FDA having recently released its draft guidance on master protocols [7] (September 2018) and enrichment designs [6] (March 2019), the number of biomarker trials will likely continue to increase. This growth in biomarker trials holds exciting promise for oncology; however, careful considerations including the biomarker tests' accuracy should be made in the planning stage. There is an additional need for increased awareness regarding the importance of accurately describing diagnostic and prognostic test characteristics [12], undertaking randomized diagnostic trials [13,14] and designing trials to simultaneously

assess test performance and treatment performance, determining how diagnostic tests can influence health outcomes directly [15].

Declaration of conflicting interests

The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

JP thanks the Mitacs Program for the Mitacs Accelerate Fellowship PhD grant (IT10703).

Funding

JP was financially supported by the Mitacs Accelerate Fellowship PhD grant. The other authors received no financial support for the research, authorship, and/or publication of this article.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.conctc.2019.100396>.

References

- [1] W.G. Feero, A.E. Guttmacher, F.S. Collins, Genomic medicine—an updated primer, *N. Engl. J. Med.* 362 (21) (2010) 2001–2011.
- [2] A.M. Bode, Z. Dong, Recent advances in precision oncology research, *NPJ Precis. Oncol.* 2 (2018) 11.
- [3] L. Schwartzberg, E.S. Kim, D. Liu, D. Schrag, Precision oncology: who, how, what, when, and when not? *Am. Soc. Clin. Oncol. Educ. Book* 37 (2017) 160–169.
- [4] M. Antoniou, A.L. Jorgensen, R. Kolamunnage-Dona, Biomarker-guided adaptive trial designs in phase II and phase III: a methodological review, *PLoS One* 11 (2) (2016) e0149803.
- [5] M. Antoniou, R. Kolamunnage-Dona, A.L. Jorgensen, Biomarker-guided non-adaptive trial designs in phase II and phase III: a methodological review, *J. Personalized Med.* 7 (1) (2017).
- [6] U.S. Department of Health and Human Services Food and Drug Administration, Enrichment strategies for clinical trials to support determination of effectiveness of human drugs and biological products guidance for industry, <https://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm332181.pdf>, (2019).
- [7] U.S. Department of Health and Human Services Food and Drug Administration, Master protocols: efficient clinical trial design strategies to expedite development of oncology drugs and biologics guidance for industry (draft guidance), <https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM621817.pdf>, (2018).
- [8] R.G. Jaeger, T.R. Halliday, On confirmatory versus exploratory research, *Herpetologica* (1998) S64–S66.
- [9] S.R. Brown, W.M. Gregory, C.J. Twelves, et al., Designing phase II trials in cancer: a systematic review and guidance, *Br. J. Canc.* 105 (2) (2011) 194–199.
- [10] B.T. Li, R. Shen, D. Buonocore, et al., Ado-trastuzumab emtansine for patients with HER2-mutant lung cancers: results from a phase II basket trial, *J. Clin. Oncol.* 36 (24) (2018) 2532–2537.
- [11] R. Simon, Optimal two-stage designs for phase II clinical trials, *Contr. Clin. Trials* 10 (1) (1989) 1–10.
- [12] J.F. Cohen, D.A. Korevaar, D.G. Altman, et al., STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration, 6 (11) (2016) e012799.
- [13] M. Roger, T. Ramsay, D. Fergusson, Diagnostic randomized controlled trials: the final frontier, *Trials* 13 (137) (2012).
- [14] S.J. Lord, L. Irwig, R.J. Simes, When is measuring sensitivity and specificity sufficient to evaluate a diagnostic test, and when do we need randomized trials? *Ann. Intern. Med.* 144 (11) (2006) 850–855.
- [15] P.M. Bossuyt, J.B. Reitsma, K. Linnet, K.G. Moons, Beyond diagnostic accuracy: the clinical utility of diagnostic tests, *Clin. Chem.* 58 (12) (2012) 1636–1643.