

How to choose a time zero for patients in external control arms

Daniel Backenroth, Janssen R&D United States

October 20, 2020

Abstract

When a sponsor carries out a single-arm trial of a novel oncology compound, it may wish to assess the efficacy of the compound via comparison of overall survival to an external control arm, constructed using patients included in some retrospective registry. If efficacy of the novel compound is compared to efficacy of physician's choice of chemotherapy, patients in the retrospective registry might qualify for inclusion in the external control arm at multiple different points in time, when they receive different chemotherapy treatments. For example, a patient might qualify at the start of their second, third and fourth lines of therapy. From the start of which line of therapy should this patient's survival be compared to survival of participants in the single-arm trial?

Some sponsors have elected to include patients in the external control arm from the last available line of therapy in the retrospective database. Another possibility is to randomly select a line of therapy for each external control arm patient from among those available. In this paper, we show, via probabilistic arguments and also via simulation based on real data, that both of these methods give rise to a bias in favor of the single-arm trial. We further show that this bias can be avoided by instead including external control arm patients multiple times in the external control arm, once for each time they receive qualifying treatment.

Introduction

Imagine that a sponsor is carrying out a single-arm trial of a novel oncology compound for patients with advanced non-small cell lung cancer. Since this is the first clinical trial for the compound, heavily pre-treated participants have been enrolled. In particular, participants are required to have been treated with prior platinum chemotherapy. Initial regulatory approval of novel oncology compounds could be based on such a single-arm trial [1, 2]. However, to provide additional context to regulators, the sponsor has decided to construct an external control arm for the trial, using data retrospectively collected from an electronic health records system [3, 4]. The sponsor will compare overall survival for patients in the external control arm to overall survival for patients in the single-arm trial, after reweighting to account for differences in patient characteristics.

One of the first steps in constructing an external control arm is to apply appropriate inclusion and exclusion requirements to the retrospective database. Here, one such inclusion criteria

might be that the patient has started physician’s choice of anticancer treatment after previous treatment with platinum chemotherapy [5]. Survival from the start of this anticancer treatment for patients in the external control arm will be compared to survival from receipt of the first dose of the novel oncology compound in the single-arm trial.

One problem in applying this criterion is that the sponsor may very well find that some patients in the retrospective database satisfy it at multiple different points in time [6, 7]. For example, consider the two patients in Table 1. Patient P2 has only one line of therapy (docetaxel) after their platinum line, and so would only qualify for inclusion in the external control arm with respect to that line of therapy. On the other hand, Patient P1 has two lines of therapy (first docetaxel plus ramucirumab, and then pemetrexed) after their platinum line, and so qualifies for inclusion in the external control arm twice. From what ”time zero” should overall survival be measured for patient P1 in the external control arm: the beginning of their second or the beginning of their third line of therapy? The timepoint selected may have a large effect on the results of the comparison to the single-arm trial, since a patient clearly survives longer from an earlier line of therapy than from a later line of therapy. Patient P1, for example, survived for nearly a year from the start of second-line treatment with docetaxel plus ramucirumab, but for less than three months from the start of third-line treatment with pemetrexed.

Table 1: Treatment history and known vital status for two fictitious patients in a retrospective database

patient id	event	date
P1	Starts Carboplatin + Vinorelbine (line 1)	2015-07-09
P1	Starts Docetaxel + Ramucirumab (line 2)	2015-10-02
P1	Starts Pemetrexed (line 3)	2016-06-21
P1	Date of death	2016-09-12
P2	Starts Carboplatin + Pembrolizumab + Pemetrexed (line 1)	2018-02-01
P2	Starts Docetaxel (line 2)	2019-01-14
P2	Administrative censor date	2019-07-31

Hernán and Robins have discussed the problem of patients in retrospective databases with multiple eligible time zeros, in the context of studies that compare two groups both extracted from the same retrospective database. They suggested that for patients with multiple eligible time zeros, there were two unbiased approaches [8]. A single eligible time, like the first eligible time or a random time, could be chosen as the time zero for each patient. Alternatively, all eligible times could be used, so that different versions of the same patient would be included in the comparison, each with a different time zero.

In this article, we describe a general model of cancer patient survival, of enrollment in clinical trials and of inclusion in retrospective database, and show that several methods of choosing a single time zero, like choosing the start of the last available line of therapy as in [6, 7], or choosing a randomly selected line of therapy, are not unbiased for an overall survival endpoint when, unlike in Hernán and Robins’ setting, comparison is made between a retrospective

database and a clinical trial (as opposed to solely within a retrospective database). We then demonstrate how the other method suggested by Hernán and Robins, in which the same patient is included multiple times in the same comparison, but with different time zeros, can be used to obtain an unbiased comparison of overall survival between a single-arm trial and patients in a retrospective database.

Model

Here we describe a general model of the survival of cancer patients, of how they enroll in a clinical trial and of how they are included in a retrospective database. As above, we take as an example population of interest non-small cell lung cancer patients who have received platinum chemotherapy and go on to receive subsequent therapy. We will see how survival of patients enrolled in a hypothetical control arm of a trial in which patients are treated with physician's choice of standard of care therapy compares to the survival of patients in an external control arm treated with physician's choice of standard of care therapy, given various choices of time zero for external control arm patients. We should choose a time zero for external control arm patients such that the survival distribution of patients in the external control arm is the same as the survival distribution of patients in this hypothetical control arm, assuming all other requirements for causal inference, e.g., no unmeasured confounders, are satisfied [9]. We call the trial control arm hypothetical since our focus here is on single-arm trials, in which there is no control arm.

Assume that each patient i in the population of interest has a time S_i , measured from some arbitrary common timepoint, for example, January 1, 2000, at which they start their first post-platinum therapy, and has N_i post-platinum lines of therapy before they die, where $N_i \geq 1$. The length of patient i 's j th post-platinum line of therapy is D_{ij} . We assume that patients die immediately after the end of their N_i th line of therapy. We further assume that the D_{ij} are independent of S_i , so that there is no time trend in survival, but that the D_{ij} can depend on N_i , so that patients with more lines of therapy can have longer (or shorter) lines of therapy. Conditional on N_i , however, we assume that D_{ij} is independent of D_{ik} for all $j \neq k$. Since we are focusing on patients receiving standard of care therapies, we assume that D_{ij} and D_{ik} , for $k > j$, do not depend on whether patient is enrolled in a clinical trial in line j . Due to censoring, N_i and some of the D_{ij} may not be observed for some patients.

Clinical trial

We assume that a trial enrolls participants from this population between times t_1 and t_2 , measured from the same common baseline time as the S_i , and that every patient in the population that finishes first-line platinum therapy during this time period, or that finishes a subsequent line of therapy during this time period and that doesn't die at the end of that line of therapy, enrolls in the trial with probability α .

Let E_i be independent random variables that are each equal to 1 with probability α and equal to 0 with probability $1 - \alpha$. Under the model above patient i will enroll in the trial if $E_i = 1$ and if there exists a k in $0 \dots N_i - 1$ such that

$$t_1 \leq S_i + \sum_{j=1}^k D_{ij} \leq t_2.$$

The sum in this expression equals 0 if $k = 0$, in which case the patient enrolls in the trial at time S_i , as their first line of therapy after platinum therapy. The line of therapy, counting starting with the first post-platinum therapy, in which patient i will enroll in the trial is

$$L_i = 1 + \min\{k \text{ such that } t_1 \leq S_i + \sum_{j=1}^k D_{ij} \leq t_2 \text{ and } k \leq N_i - 1\}$$

The survival of patient i from enrollment in the trial until death is $Y_i = \sum_{j=L_i}^{N_i} D_{ij}$. For some patients Y_i may be censored and not known at the time of comparison to the external control arm.

Retrospective registry

We assume that a retrospective registry of cancer patients includes, with probability β , any patient from this population for whom $S_i < t_c$, where t_c is the data cutoff date for the registry. That is, if the patient starts their line of treatment after platinum chemotherapy after the date t_c , then they won't be included in the registry. Let F_i be independent random variables that are each equal to 1 with probability β and equal to 0 with probability $1 - \beta$. We assume that the registry includes complete information until time t_c on the treatment history of any patient with $F_i = 1$ and $S_i < t_c$. Any events occurring after time t_c are censored. The number of post-platinum lines of therapy observed for a patient in the registry depends on the relationship between S_i , t_c , N_i and the D_{ij} . Letting M_i be the number of post-platinum lines of therapy whose start time is observed for patient i in the retrospective database, which may be lower than N_i due to censoring, we have that

$$M_i = 1 + \max\{k \text{ such that } S_i + \sum_{j=1}^k D_{ij} \leq t_c \text{ and } k \leq N_i - 1\}.$$

The survival of patient i from the start of their k th post-platinum line, $1 \leq k \leq N_i$, is $W_{ik} = \sum_{j=k}^{N_i} D_{ij}$, although this survival may be censored.

Overall survival observed in clinical trial

We want to check if $P(Y_i \geq y | L_i = l, E_i = 1)$, the survival function for patients enrolling in the trial in line setting l , is the same as $P\left(\sum_{j=l}^{N_i} D_{ij} \geq y | N_i \geq l\right)$, the marginal survival function for patients in the population, from that line setting, regardless of enrollment in the trial. We have

$$\begin{aligned}
 P(Y_i \geq y | L_i = l, E_i = 1) &= P\left(\sum_{j=l}^{N_i} D_{ij} \geq y | S_i + \sum_{j=1}^{l-2} D_{ij} < t_1, t_1 \leq S_i + \sum_{j=1}^{l-1} D_{ij} \leq t_2, N_i \geq l, E_i = 1\right) \\
 &= P\left(\sum_{j=l}^{N_i} D_{ij} \geq y | N_i \geq l, E_i = 1\right) \\
 &= P\left(\sum_{j=l}^{N_i} D_{ij} \geq y | N_i \geq l\right) \tag{1}
 \end{aligned}$$

The first line above follows since a patient will enroll in the trial in line setting l only if the patient's $l - 1$ th post-platinum line is the first line of that patient to end in the interval between t_1 and t_2 (if $l = 1$ then the first inequality conditioned on is ignored). The second line follows from mutual independence of the D_{ij} and from mutual independence of the D_{ij} and the S_i , and the third line follows from independence of the D_{ij} and E_i . Therefore, under the model posed above, the survival of patients enrolling in the trial in a particular line setting l , i.e., those with $L_i = l$, is the same as the marginal survival of patients in the population from the l th line.

Conditioning on $N_i = n$, we can also write this marginal probability as follows:

$$P\left(\sum_{j=l}^{N_i} D_{ij} \geq y | N_i \geq l\right) = \sum_{n=l}^{\infty} P\left(\sum_{j=l}^n D_{ij} \geq y | N_i = n\right) P(N_i = n) \tag{2}$$

Note that equations 1 and 2 would not hold if the D_{ij} were not conditionally independent given N_i . Here we sum to infinity, although in practice the N_i are bounded. In the appendix, we show, assuming that the times at which patients start post-platinum therapy S_i are uniformly distributed, that if the lengths of the D_{ij} are positively correlated within a patient then patients with better survival may be over-represented in a clinical trial.

Overall survival observed in retrospective database, choosing one eligible line per patient

One potential way to set up an external control arm is to measure survival for each patient in the retrospective database from one post-platinum line of therapy chosen using some rule from the available post-platinum lines of therapy for that patient. Let R_i be the selected post-platinum line of therapy from which survival is measured for patient i . The selection of R_i could depend on M_i . For example, one line might be chosen randomly from among the available lines of therapy in a database for a patient, or the last available line of therapy might be chosen for a patient. Alternatively, the selection of R_i could be independent of M_i . For example, the first line could always be chosen, $R_i = 1$. As another example, R_i could equal 1 with probability 0.5 and 2 with probability 0.5, irrespective of M_i . With this

scheme, a patient with $R_i > M_i$ would be excluded from the external control arm, and so sample size would be affected.

We first assume that we have complete capture for each patient in the retrospective registry, i.e., N_i is observed for each patient in the registry, so that $N_i = M_i$. The survival for patients from line setting l in an external control arm so constructed, $P(W_{il} \geq y | R_i = l)$, is, conditioning on N_i and then applying Bayes' rule:

$$P(W_{il} \geq y | R_i = l) = \sum_{n=l}^{\infty} P(W_{il} \geq y | R_i = l, N_i = n) P(N_i = n | R_i = l)$$

$$= \sum_{n=l}^{\infty} P(W_{il} \geq y | N_i = n) P(R_i = l | N_i = n) \frac{P(N_i = n)}{P(R_i = l)} \quad (3)$$

$$= \sum_{n=l}^{\infty} P(W_{il} \geq y | N_i = n) \frac{P(R_i = l | N_i = n) P(N_i = n)}{\sum_{m=l}^{\infty} P(R_i = l | N_i = m) P(N_i = m)} \quad (4)$$

Here we sum N_i from l to ∞ , but the N_i in practice are bounded above. Note that in the second line we assume that W_{il} is independent of R_i given N_i , i.e., that the selection of R_i may depend on how many lines of therapy a patient has, but not on other features of the survival history of the patient.

Let us consider some rules for selecting a post-platinum line of therapy for each patient in the external control arm. First we assume that we select a line uniformly at random for each patient from the available lines of therapy, that is, $P(R_i = l | N_i = n)$ for $l \leq n$ equals $1/n$. Then equation 4 equals

$$\sum_{n=l}^{\infty} P\left(\sum_{j=l}^{N_i} D_{ij} \geq y | N_i = n\right) \frac{P(N_i = n)/n}{\sum_{m=l}^{\infty} P(N_i = m)/m} \quad (5)$$

so we see that, by comparison to the weighted sum for the marginal survival probability observed in the trial (equation 2), probabilities for patients with more lines of therapy, and hence longer survival, are given smaller weights in this expression, leading to bias if this method is adopted.

In fact, by inspecting equations 2 and 3, we find that in general, in order to avoid bias, the selection of R_i for each patient must not depend on N_i . So, for example, choosing the last line for each external control arm will lead to bias. In this case $P(R_i = l | N_i = n) = 1$ for $l = n$ and 0 otherwise, so that equation 3 becomes $P(W_{il} \geq y | N_i = l)$, and we are only estimating survival from line setting n using patients who died immediately after that line setting.

By contrast, always choosing the first line for each patient, $P(R_i = 1) = 1$, will not lead to bias, but then no external control arm patients will be able to be compared to trial patients who received active comparator in a line after their first post-platinum line of therapy. Choosing a line of therapy at random for each patient where the random selection does not

depend on N_i will avoid bias, but this choice will mean that many patients will be excluded from the external control if they did not survive until the randomly chosen line of therapy.

Before, we ignored the fact that we did not collect data for patients in the registry after the data cutoff date t_c . Now we take this into account. First, we consider selection rules where R_i is independent of M_i and of the survival history of the patient. For example, assume we let $R_i = 1$ with probability $\frac{1}{2}$ and $R_i = 2$ with probability $\frac{1}{2}$. We want to check that $P(W_{il} \geq y | R_i = l, M_i \geq l, F_i = 1)$, the survival of external control arm patients selected from the registry, is the same as $P(\sum_{j=l}^{N_i} D_{ij} \geq y | N_i \geq l)$, the marginal survival of patients from that line setting. If so, then we can use such selection rules to select an external control arm whose survival distribution can be compared without bias to the survival distribution of patients in the trial.

We have:

$$\begin{aligned}
 P(W_{il} \geq y | R_i = l, M_i \geq l, F_i = 1) &= P\left(\sum_{j=l}^{N_i} D_{ij} \geq y | R_i = l, M_i \geq l, F_i = 1\right) \\
 &= P\left(\sum_{j=l}^{N_i} D_{ij} \geq y | S_i + \sum_{j=1}^{l-1} D_{ij} \leq t_c, N_i \geq l\right) \\
 &= P\left(\sum_{j=1}^{N_i} D_{ij} \geq y | N_i \geq l\right) \tag{6}
 \end{aligned}$$

The second line here follows from the independence of F_i and R_i with all other variables, and the third line follows from the mutual independence of the D_{ij} and from mutual independence of the D_{ij} and S_i .

The appendix discussed the the bias arising from a selection rule that selects a line uniformly at random for each patient from the available post-platinum lines of therapy of that patient, in this general case where only M_i line starts are observed for each patient.

Overall survival observed in retrospective database, choosing all eligible lines per patient

Instead of choosing a single eligible line for each patient, Hernán and Robins suggested that all eligible timepoints could be used. Applying this idea to our context, the comparison of external control arm patients to trial patients could be stratified by line, and an external control arm patient could be included in multiple strata, once for each line they had after platinum therapy. In each strata, the patient's survival would be measured from the appropriate line of therapy and compared to patients enrolling in the clinical trial in that line setting. By equation 6, this would be an unbiased comparison, since the selection probabilities for patients in the external control arm would be $P(R_i = l | N_i = n) = 1$ regardless of the value of n . However, the correlations between survival times for patients included in multiple strata need to be accounted for. Here we account for these correlations using the grouped

approximate jackknife estimate of the variance described in [10], which is implemented in standard statistical software [11].

Simulation study

As shown above, if we choose a line of therapy at random for each patient in a retrospective database, but the random selection does not depend on M_i , the number of lines observed for that patient, then our estimate of survival from each line setting will be unbiased. We can use this method to carry out a simulation in which we construct repeated single-arm trials, along with external control arms. Here the patients in both arms receive the same therapy and come from the same source population, so that if the external control arm is constructed correctly outcomes should be the same as in the single-arm trial.

To carry out this simulation study, we use the nationwide Flatiron Health electronic health record (EHR)-derived de-identified database. The Flatiron Health database is a longitudinal database, comprising de-identified patient-level structured and unstructured data, curated via technology-enabled abstraction. During the study period, the de-identified data originated from approximately 280 cancer clinics (800 sites of care). The study included 15,243 patients diagnosed with advanced non-small cell lung cancer from January 2011 to April 2020. The majority of patients in the database originate from community oncology settings; relative community/academic proportions may vary depending on study cohort. Lines of therapy in this database are oncologist-defined and rule-based. The data are de-identified and subject to obligations to prevent re-identification and protect patient confidentiality. Flatiron Health, Inc. did not participate in the analysis of this data.

We imagine we are using this dataset to construct an external control arm for a new compound being administered to lung cancer patients who have previously received platinum therapy. We allow patients in the external control arm to receive any therapy, i.e., we are comparing to physician's choice of therapy, but we restrict to patients receiving therapy in the 5th or earlier line. There are 15,243 patients who meet these criteria in one or more line. 37% of these patients meet these criteria in only one treatment line; 30% meet them in two different lines and the remaining patients meet these criteria in three or four different lines.

We repeatedly simulate both a single-arm trial as well as an external control arm. Note that here the therapy is the same in the single-arm trial as in the external control arm, so that the survival distribution in both should be the same. To do this, for each simulation run we start by splitting our overall sample into two parts. From one part we select patients for the single-arm trial, using fixed probabilities to select patients in different line settings, so that on average trial patients will be evenly distributed across line numbers 2, 3, 4 and 5. From the other part we randomly select patients for the external control arm. We simulate single-arm trials and external control arms with 40, 160 and 640 patients.

We use three different methods to select lines for external control arm patients to compare to patients in the simulated single-arm trial: random selection of a line uniformly at random from the observed post-platinum lines of the patient, selection of the patient's last line, and inclusion of all lines. We compare the three external control arms so constructed to the trial

patients by calculating a hazard ratio (with the trial patients as the reference group), and its associated p-value, stratifying by line in a Cox proportional hazards model. We calculate bias as the mean hazard ratio (on the logarithmic scale) comparing the external control arm patients to the trial patients. In the absence of this bias this log hazard ratio should be equal to 0. We use 5,000 replicates per sample size.

Table 2 shows the bias that results from random selection of lines or using the last line for external control arm patients. This bias alone results in a hazard ratio of between 1.12 and 1.15 comparing external control arm patients to single-arm patients; this is an anti-conservative bias, as it will tend to exaggerate the effect of the therapy provided to single-arm patients. Table 2 also illustrates the lack of bias from the method in which all qualifying lines for all patients are used, with mean bias close to 0. Note that bias results from using random selection as the random selection of a line for an external control arm patient depends on how many lines a patient has in the database. Table 2 also shows the inflation of Type I error that results when using the naive variance estimator while including all lines for each external control arm patient; Type I error is more than doubled compared to the nominal 5% rate. For 640 patients, type I error is adequately controlled using the grouped approximate jackknife estimate, when all lines are used for each external control arm patients. Type I error is inflated only for lower sample sizes.

Table 2: Simulation results. Mean bias is defined as the mean hazard ratio on the logarithmic scale, comparing simulated external control arm patients to simulated single-arm trial patients. In the absence of bias this log hazard ratio should equal 0. Testing was carried out at the 5% level so Type I error above 5% indicates inflation. For the all lines method, all available lines for external control arm patients were compared to the single-arm patients. For the last available line method, survival was measured for external control arm patients from the last line of therapy observed in the database. For the randomly selected line method, survival was measured for each external control arm patient from a line of therapy selected uniformly at random from among the available lines of therapy for that patient.

type	number per arm	Type I error	Mean bias
All lines, approximate jackknife variance	40	0.078	-0.006
All lines, approximate jackknife variance	160	0.062	-0.007
All lines, approximate jackknife variance	640	0.051	-0.001
All lines, naive variance	40	0.101	-0.006
All lines, naive variance	160	0.109	-0.007
All lines, naive variance	640	0.108	-0.001
Last available line	40	0.069	0.149
Last available line	160	0.135	0.136
Last available line	640	0.390	0.132
Randomly selected line	40	0.064	0.134
Randomly selected line	160	0.120	0.123
Randomly selected line	640	0.320	0.121

Conclusion

What should be done when a patient in an external control arm for an oncology trial satisfies inclusion and exclusion criteria at the start of more than one line of therapy? Here we have shown that some methods of selecting a single line of therapy for each external control arm patient, like selecting a line of therapy uniformly at random, may lead to substantial bias, exaggerating the effect of the treatment provided to single-arm trial patients. These methods should therefore be avoided. By contrast, including all eligible lines for external control arm patients avoids this bias. We have demonstrated this via a model of the survival of cancer patients and their recruitment into a trial or inclusion in a retrospective database, as well as via a simulation study using data from a retrospective database. We have also demonstrated how to account for inflation of Type I error arising when more than one outcome for the same patient is included. The approximate jackknife estimator whose performance is evaluated here adequately controls Type I error, especially at larger sample sizes. Further work could assess whether the bootstrap better controls Type I error at smaller sample sizes.

References

- [1] Sharyl J. Nass, Mace L. Rothenberg, Rebecca Pentz, Hedvig Hricak, Amy Abernethy, Kenneth Anderson, Amanda Wagner Gee, R. Donald Harvey, Steven Piantadosi, Monica M. Bertagnolli, Deborah Schrag, and Richard L. Schilsky. Accelerating anticancer drug development —opportunities and trade-offs. *Nature Reviews Clinical Oncology*, 15(12):777–786, 2018.
- [2] John R. Johnson, Yang-Min Ning, Ann Farrell, Robert Justice, Patricia Keegan, and Richard Pazdur. Accelerated Approval of Oncology Products: The Food and Drug Administration Experience. *JNCI: Journal of the National Cancer Institute*, 103(8):636–644, 2011.
- [3] Elodie Baumfeld Andre, Robert Reynolds, Patrick Caubel, Laurent Azoulay, and Nancy A. Dreyer. Trial designs using real-world data: The changing landscape of the regulatory approval process. *Pharmacoepidemiology and Drug Safety*, 2019.
- [4] Bhakti Arondekar, Rachel Bhak, Maral DerSarkissian, Lynn Huynh, Kelsey Wang, Eric Davis, Bryon Wornson, and Mei Sheng Duh. Role of real-world evidence for oncology product registration in the United States: A review of approvals by the U.S. Food and Drug Administration from 2015 to 2019. *Journal of Clinical Oncology*, 38(15_suppl):e14130–e14130, 2020.
- [5] ASCENT-Study of Sacituzumab Govitecan in Refractory/Relapsed Triple-Negative Breast Cancer. <https://clinicaltrials.gov/ct2/show/NCT02574455>, note = Accessed: 2020-09-23.
- [6] Sundar Jagannath, Yi Lin, Hartmut Goldschmidt, Donna Ellen Reece, Ajay K. Nooka, Paula Rodríguez Otero, Kosei Matsue, Nina Shah, Larry D. Anderson, Jr, Kimberly Wilson, Arlene S. Swern, Faiza Zafar, Amit B. Agarwal, and David Samuel DiCapua

Siegel. KarMMa-RW: A study of real-world treatment patterns in heavily pretreated patients with relapsed and refractory multiple myeloma (RRMM) and comparison of outcomes to KarMMa. In *ASCO Virtual Scientific Program*, 2020.

- [7] Non-interventional study protocol: Retrospective cohort of relapsing/refractory multiple myeloma (RRMM) patients, treated in terms of clinical practice, who have received at least three prior lines of anti-myeloma therapy to be compared with patients enrolled in GSK’s clinical trial NCT03525678 of GSK2857916 medicine (BCMA). https://www.pei.de/SharedDocs/Downloads/DE/awb/nis-0401-0500/0492-beoplan.pdf?__blob=publicationFile&v=4, note = Accessed: 2020-07-20.
- [8] Miguel A. Hernán and James M. Robins. Using big data to emulate a target trial when a randomized trial is not available. *American Journal of Epidemiology*, 183(8):758–764, 03 2016.
- [9] Miguel A. Hernán and James M. Robins. *Causal inference: What if*. Chapman and Hall/CRC, Boca Raton, 2020.
- [10] Terry M. Therneau. Extending the Cox model. Technical Report 58, Mayo Clinic, Rochester, Minnesota, 1996.
- [11] Terry M Therneau. *A Package for Survival Analysis in R*, 2020. R package version 3.2-3.

1 Appendix

1.1 Overall survival observed in clinical trial if D_{ij} are positively correlated

Assume that the D_{ij} are positively correlated even when conditioning on N_i , the total number of lines of therapy, so that a patient that has one longer than average line of therapy is also likely to have other longer than average lines of therapy. As an example of the selection bias that can occur as a result, let us consider the probability that a patient who has n post-platinum lines of therapy will be selected into the trial at the start of their 2nd post-platinum line.

This probability is $P(S_i < t_1, S_i + D_{i1} > t_1, S_i + D_{i1} < t_2, E_i = 1 | N_i = n)$. Let us assume that S_i , the time of starting first post-platinum therapy, is uniformly distributed conditional on N_i , i.e., that treatment patterns have not changed recently. Then this probability is proportional to $P(t_1 - D_{i1} < S_i < \min(t_1, t_2 - D_{i1}) | N_i = n) = \min(t_1, t_2 - D_{i1}) - (t_1 - D_{i1})$. If $D_{i1} \geq t_2 - t_1$, the length of the enrollment period, then this expression is equal to a constant, $t_2 - t_1$. Otherwise, it equals D_{i1} . We therefore see that the probability of selection in the trial is lower for patients with shorter D_{i1} than for patients with longer D_{i1} . Since the D_{ij} are positively correlated, patients in the trial will then have longer post-enrollment survival than patients in the general population, conditioning on N_i .

1.2 Selecting a line uniformly at random for each patient with censoring

When N_i is not observed for some patients in the retrospective database due to administrative censoring, then survival may be measured for each patient in the retrospective database from a post-platinum line of therapy chosen using a rule that depends on M_i , the number of post-platinum lines of therapy observed for a patient in the registry. The survival for patients from line setting l in an external control arm so constructed is, conditioning on both N_i and M_i and applying Bayes' rule:

$$\begin{aligned} P(W_{il} \geq y | R_i = l) &= \sum_{n=l}^{\infty} \sum_{m=l}^n P(W_{il} \geq y | R_i = l, N_i = n, M_i = m) P(N_i = n, M_i = m | R_i = l) \\ &= \sum_{n=l}^{\infty} \sum_{m=l}^n P(W_{il} \geq y | R_i = l, N_i = n, M_i = m) \\ &\quad \times P(R_i = l | N_i = n, M_i = m) \frac{P(N_i = n, M_i = m)}{P(R_i = l)} \end{aligned}$$

If we assume that a post-platinum line of therapy is chosen uniformly at random from among the M_i post-platinum lines of therapy for a patient, then this expression becomes

$$P(W_{il} \geq y | R_i = l) = \sum_{n=l}^{\infty} \sum_{m=l}^n P(W_{il} \geq y | N_i = n, M_i = m) \frac{P(N_i = n, M_i = m)/m}{P(R_i = l)}.$$

Here W_{il} is independent of R_i conditional on M_i since the selection of a line of therapy depends only on M_i and not on other features of the survival of the patient.

Let us assume, to keep things simple, that N_i can only equal 1 or 2, i.e., that patients can have at most 2 post-platinum lines of therapy. Had N_i been observed in this simple case and R_i had been uniformly selected from among $1, \dots, N_i$, then survival from the first post-platinum line for the external control arm would have been estimated (see equation 5) as

$$P(W_{i1} > y | R_i = 1) = \frac{P(W_{i1} > y | N_i = 1)P(N_i = 1) + P(W_{i1} > y | N_i = 2)P(N_i = 2)/2}{P(N_i = 1) + P(N_i = 2)/2} \quad (7)$$

This is a biased estimate, compared to the survival probabilities as measured in a single arm trial (equation 2), due to the downweighting of the survival contribution from patients who actually survive for two lines of therapy.

In the general case where N_i is not observed due to administrative censoring, then instead survival from the first post-platinum line would be estimated as

$$P(W_{i1} > y | R_i = 1) = \left(P(W_{i1} > y | N_i = 1)P(N_i = 1) + P(W_{i1} > y | N_i = 2, M_i = 1)P(N_i = 2, M_i = 1) + \frac{1}{2}P(W_{i1} > y | N_i = 2, M_i = 2)P(N_i = 2, M_i = 2) \right) / \left(P(N_i = 1) + P(N_i = 2, M_i = 1) + \frac{1}{2}P(N_i = 2, M_i = 2) \right)$$

In this general case, we see that more weight is given in this weighted sum to survival from line 1 of those patients who actually survived for 2 lines, as compared to the biased estimate in equation 7. In particular, the contribution from patients who survive for 2 lines but whose start of line 2 was administratively censored is not downweighted at all. These patients may actually be expected to have better survival than those patients who survived for 2 lines but whose start of line 2 was not administratively censored, since patients with shorter first lines of therapy are less likely to be administratively censored. Therefore in this general case less bias is to be expected than in the case where N_i is observed for all patients.

Although bias arising from uniform random selection among available lines of therapy is attenuated as a result of administrative censoring, as compared to what would be expected from equation 5, it is important to note that most patients in a retrospective database may have observed dates of death, although the proportion will depend on how many years of data the retrospective database includes and the survival of patients in the database.

2 Data availability

The data that support the findings of this study have been originated by Flatiron Health, Inc. These de-identified data may be made available upon request, and are subject to a license agreement with Flatiron Health; interested researchers should contact DataAccess@flatiron.com to determine licensing terms.