Estimands in hematologic oncology trials

Steven Sun1 | Hans-Jochen Weber2 | Emily Butler3 | Kaspar Rufibach4 | Satrajit Roychoudhury5

1Statistics and Decision Science, Janssen Research and Development LLC, Raritan, New Jersey, USA
2Clinical Development and Analytics, Novartis Pharma AG, Basel, Switzerland
3Biostatistics, GlaxoSmithKline, Collegeville, Pennsylvania, USA
4Methods, Collaboration, and Outreach Group, Product Development Data Sciences, F. Hoffmann-La Roche, Basel, Switzerland
5Statistical Research and Innovation, Global Biometrics and Data Management, Pfizer Inc., New York, New York, USA

Correspondence
Steven Sun, Statistics and Decision Science, Janssen R&D, 920 Route 202, Raritan, NJ 08502, USA.
Email: ssun@its.jnj.com

Abstract
The estimand framework included in the addendum to the ICH E9 guideline facilitates discussions to ensure alignment between the key question of interest, the analysis, and interpretation. Therapeutic knowledge and drug mechanism play a crucial role in determining the strategy and defining the estimand for clinical trial designs. Clinical trials in patients with hematological malignancies often present unique challenges for trial design due to complexity of treatment options and existence of potential curative but highly risky procedures, for example, stem cell transplant or treatment sequence across different phases (induction, consolidation, maintenance). Here, we illustrate how to apply the estimand framework in hematological clinical trials and how the estimand framework can address potential difficulties in trial result interpretation. This paper is a result of a cross-industry collaboration to connect the International Conference on Harmonisation (ICH) E9 addendum concepts to applications. Three randomized phase 3 trials will be used to consider common challenges including intercurrent events in hematologic oncology trials to illustrate different scientific questions and the consequences of the estimand choice for trial design, data collection, analysis, and interpretation. Template language for describing estimand in both study protocols and statistical analysis plans is suggested for statisticians’ reference.

KEYWORDS
estimand, hematology clinical trials, ICH E9 addendum, intercurrent events

1 INTRODUCTION

The key questions of interest in a clinical trial vary from one stakeholder to another. The current practice of defining trial objectives rather vaguely has led to various challenges in interactions with different stakeholders, including regulators and payers. To improve transparency and ensure alignment between trials objectives, data collection and statistical approaches, it appears necessary to go back to the first principles and to clearly define what is to be estimated.

The addendum to the ICH E9 guideline was released in 2019 to introduce the estimand framework. This framework targets to align the clinical trial objectives with the design and statistical analyses by providing detailed definitions of the quantity of scientific interest (“estimand”). Besides estimands, the addendum also discusses the role of sensitivity and supplementary analyses. Therapeutic knowledge and drug mechanism play a crucial role in determining the estimand of interest and the design.1

In recent years, many clinical trials have been initiated in hematologic malignancies due to an enhanced understanding of cancer pathologies.2 Hematological malignancies such as leukemia, multiple myeloma and lymphoma
represent a subset of indications in hematologic oncology (referred to as “hematology” for simplicity below). Clinical trials in hematologic malignancies are characterized by specific definition of endpoints, data collection, magnitude of clinically meaningful treatment effect, and interpretation of treatment effect. Moreover, standard therapies are often following a sequence of treatment (e.g., induction, consolidation, maintenance) and offering stem cell transplantation can be an important part of such a treatment strategy. This can lead to challenges when defining the treatment attribute of an estimand. \(^3\) \(^4\) Time-to-event (TTE) endpoints are mainstays of late phase hematology clinical trials. Though, overall survival (OS) is considered the gold standard for determining treatment efficacy in oncology trials, but OS is traditionally defined under treatment policy, which may not address the underlying question of treatment difference if no subsequent therapies are available. \(^5\) It can also be challenging to assess because of long trial durations. Therefore, other clinical endpoints like progression-free survival (PFS), event-free survival (EFS), time to progression and response rate (RR) may be appropriate endpoints for the regulatory approval of treatments. \(^6\) \(^8\) Some statistical analyses may be more intuitive and easier to interpret than others, depending on the key question of interest, trial type, patient population, and duration of follow-up. Another typical feature for a hematologic malignancy is that the broad population defined in a clinical trial for the same type of cancer is often heterogeneous in the sense that patients could have very different types of genomic characteristics. For example, some patients with diffuse large B cell lymphoma (DLBCL) may be curable while others may die quickly irrespective of any treatment options available so far. \(^9\)

In this paper, we present the utility of the estimand framework for typical hematology trials and provide practical recommendations. The estimand framework lends itself to a more transparent way of specifying each objective of a trial and ensuring alignment with the selected estimator/analysis method so that the most appropriate methods for primary and sensitivity analyses are identified and applied. In this paper we discuss ways to incorporate the estimand framework, especially strategies to common intercurrent events when designing a clinical trial for hematologic malignancies. Issues relevant to defining estimands are discussed. On the side of the estimator, considerations are given regarding the trial design to minimize dropouts and missing values. In addition, we present real examples to illustrate the application of different estimand strategies (Table 1).

The rest of the paper is structured as follows. In Section 2, we introduce three typical case trials to illustrate the key aspects of the estimand framework for hematology trials. Sections 3 and 4 focus on the strategies for different intercurrent events and relevant sensitivity analyses and supplementary analyses. Finally, the impact of the estimand framework on trial design, data collection and data analysis are discussed in Section 5. Section 6 concludes with a discussion. Sample language for protocol and statistical analysis plan (SAP) are provided in Data S1.

## 2 | MOTIVATING EXAMPLES

We provide three case studies to demonstrate the utility of the estimand framework in confirmatory hematology trials. These examples consist of randomized Phase 3 trials with time to event endpoint comparing a novel treatment with standard of care (SOC). They include the disease areas non-Hodgkin’s lymphoma, multiple myeloma, and acute myeloid leukemia. The purpose of this exercise is to demonstrate the application of the estimand framework in hematology to practitioners. We have no intention to comment or judge the clinical activity of the drugs involved or related regulatory decisions. Further details of the treatment, population, primary endpoint, and design are shown in Table 2 and will be used in later sections for introducing appropriate estimands.

Further details of the three trial designs are schematically shown in Figure 1. For all three studies, the design has two phases: induction and maintenance. In addition to that, the multiple myeloma and the RATIFY trial also have a consolidation phase. During the induction (or consolidation) phase, patients are treated aggressively with the experimental drug and/or SOC. This is followed by maintenance therapy until disease progression or unacceptable toxicity. For all the trials, the comparisons are made with a primary endpoint which is measured over all phases following a traditional intent to treat (ITT) principle. However, the design cannot address other important clinical questions such as the benefit of treatment in each phase due to the occurrence of different intercurrent events (e.g., start of maintenance therapy, stem cell transplantation [ASCT or HSCT], possible response or remission which decides the maintenance therapy etc.). Key assumption like proportional hazard between two treatment groups is questionable in presence of these intercurrent events (e.g., ASCT may change the disease course completely). Other challenges include interest in additional patient populations (marginal-zone lymphoma, MZL in Gallium trial, for which the primary interest is in follicular lymphoma, FL patients), discrepancies between investigator and independent reviewer regarding the assessment of progression, and assessment of secondary
endpoints (e.g., event-free survival, EFS). In the next section we show how the five estimand attributes can be used to address these issues and answer appropriate clinical questions of interest.

For all the examples above, intent-to-treat (ITT) analyses were planned for the primary analysis. The clinical meaningfulness of such estimates needs to be assessed as this does not consider that some patients take new anti-cancer therapy during the trial. This may convolute the overall biological effect of the treatment. For example, consider a situation where most patients on placebo switch to a new anti-cancer therapy but only a few on the active drug switch. In this case, the ITT analysis essentially compares the new drug with another anti-cancer therapy, rather than with placebo. Moreover, a series of sensitivity analysis are often pre-specified in the analysis plan to show robustness of the treatment effect. Often, these analyses lack the description of the underlying estimand which creates issues with clinical interpretation. In the next section we show the utility of the estimand framework to address these challenges.

### 3 | APPLICATION OF THE ESTIMAND FRAMEWORK IN HEMATOLOGIC ONCOLOGY

In this section we discuss how the key questions in hematological studies can be viewed in the estimand framework. Based on the ICH E9 addendum it is defined through the five attributes viz, treatment, population, variable,
intercurrent event, and population-level summary. Each intercurrent event and associated strategy needs to be clearly stated. Intercurrent events are clinical events that occur after treatment initiation that affect either the interpretation or the existence of the measurements associated with the clinical question of interest. Some potential examples of intercurrent events in hematology trials are discontinuation of treatment, start of maintenance therapy, or treatment switching.

**TABLE 2** Attributes of estimands as per ICH E9 addendum for three case studies

| Scientific question | Gallium study | Multiple myeloma study | RATIFY study |
|---------------------|---------------|------------------------|--------------|
| Will the addition of Obinutuzumab to FL patients treatment strategy prolong the time to death and progression regardless of new anti-lymphoma treatments prior to experiencing a PFS event | Will the addition of new drug to VRd to a treatment strategy consisting of induction, ASCT, consolidation and maintenance prolong the time to death and progression compared to VRd alone if patients with new therapy would have had the same subsequent event hazards as patients without new therapy, had they not taken new therapy | Will the addition of Midostaurin to newly diagnosed AML treatment strategy prolong the time to death regardless of new therapies and SCT |
| Primary objective | To evaluate the efficacy of G-chemo followed by G-maintenance therapy compared with R-chemo followed by R-maintenance therapy in patients with previously untreated advanced FL, as measured by investigator-assessed PFS. | To evaluate the efficacy of the new drug to a standard care VRd as a treatment sequence in patients with newly diagnosed multiple myeloma who are transplant eligible in terms of PFS using algorithm. | To evaluate benefit of Midostaurin in addition to the standard care in prolonging overall survival compared with the standard care in patients newly diagnosed AML treatment strategy irrespective of any supportive care and subsequent therapy use. |
| Treatment | Patients receive 6 or 8 induction cycles of induction chemoimmunotherapy. Responding patients receive single agent therapy during maintenance therapy for up to 2 years | Patients receive four cycles induction therapy. Responding patients receive ASCT followed by two cycles of consolidation therapy. Thereafter patients are entering maintenance therapy | Patients receive one cycle induction therapy. If no remission, 2nd induction cycle is given. Eligible patients receive SCT. Otherwise up to four cycles of consolidation. After consolidation patients receive 12 cycles single-agent maintenance treatment |
| Population | FL patients as defined by protocol eligibility criteria | MM patients as defined by protocol eligibility criteria | First-line AML patients as defined by eligibility criteria |
| Variable | PFS defined as time from randomization to PD as per investigator assessment or death, whichever is earlier | PFS by algorithm based on lab data and clinical data | OS |
| Intercurrent events and strategies | 1. New anti-cancer therapy prior to PFS event | Treatment Policy: intercurrent event is ignored, PFS time irrespective of start of new therapies is considered | Hypothetical: Patients who received anti-cancer therapy are assumed to have the same risk as those who did not receive subsequent anti-cancer therapy |
| | Treatment Policy: intercurrent event is ignored |
| | Treatment Policy: intercurrent event is ignored |
| | Treatment Policy: intercurrent event is ignored |
| Population-level summary | Hazard ratio | Hazard ratio | Hazard ratio |

---

aThe occurrence of the intercurrent event is considered irrelevant in defining the treatment effect of interest.

bHazard ratio calculated using stratified cox regression with treatment as the only covariate.
The treatment effect is quantified using a population-level summary. There could be multiple ways of describing the same estimand. For example, a composite strategy for an intercurrent event of subsequent therapy together with the endpoint of PFS is the same as treatment policy strategy for an intercurrent event of subsequent therapy when the endpoint is defined as EFS, which is time interval from randomization to the earliest of progressive disease (PD), subsequently therapy or death.

All these case studies in Section 2 are large phase 3, randomized, controlled trials in patient populations with different hematologic indications and represent typical trial designs in hematology drug development. For each example, we now describe the primary scientific question and the corresponding five estimand attributes. These examples help practitioners to understand how to explain the key question of interest using the estimand framework in the study protocol and analysis plan. All the details are provided in Table 2. Note that these trials were designed and executed prior to the release of the ICH E9 addendum. Therefore, this exercise is performed retrospectively for illustration purposes only.

**FIGURE 1** Schematic overview of the study designs for three examples
The case studies represent typical treatment approaches in hematologic indications consisting of different treatment parts. The experimental treatment is often added to the SOC treatment. Trial designs comparing treatment strategies (combination or treatment sequence) will allow proper assessment of the whole strategy. However, it is extremely difficult if not impossible to estimate the contribution of an individual component of the strategy unless 2nd randomization or multiple trials are implemented. Health authorities might request such information if some parts of the treatment strategy are not well established like the maintenance treatment in AML or myeloma. With the estimand framework, there is no ambiguity about what treatment effect the study is designed to establish.

The same considerations apply to stem cell transplantation (SCT). In many hematologic diseases, SCT is part of the standard treatment options. In AML, SCT improves the chance of cure but relates also to specific risks like graft-vs-host disease or transplant-related mortality. As such, SCT might impact the results for OS or PFS. The question is whether we need to assess the impact of SCT on OS or PFS. However, the receiving SCT might be a result of the experimental therapy and thus, assessing the contribution of the elements of a treatment strategy might be challenging or even not be feasible. Assessing the effect of the whole treatment strategy integrating all its components might be the most relevant approach for the patients. The estimand framework will facilitate these discussions with the health authorities prior to starting a study. More specifically, sponsor and the health authority can agree on what can be estimated and what cannot be estimated at design stage.

The estimands described in Table 2 clearly define what is to be estimated and how to estimate treatment benefit based on the scientific questions of interest. The estimand framework ensures consistency of the choice of statistical analysis with the scientific question of interest. This avoids misalignment between the trial objectives, analysis, and interpretation of result. Different strategies used to address intercurrent event of subsequent therapy address different clinical questions and this becomes clearer when embedding these clinical questions in the estimand framework. Different stakeholders will likely be interested in different scientific questions. For example, health authorities may be interested in the OS benefit irrespective of subsequent anti-cancer therapy while HTA (Health Technology Assessment) is interested in the OS benefit if no subsequent anticancer is allowed.

Under the estimand framework, ambiguity of interpretation of trial results will be eliminated or minimized. Each scientific question of importance can be addressed by the mutual agreement of estimand description between trial sponsor and the targeted stakeholder (health authorities, HTA, patients, etc.) for assessing the trial results.

4 SENSITIVITY ANALYSIS VERSUS SUPPLEMENTARY ANALYSIS

Description of estimand as illustrated in case studies in Section 3 above includes the main estimator or summary of treatment effect. The ICH E9 addendum also provides guideline to assess the robustness of treatment effect. An estimand is associated with one specific scientific question of interest. However, there could be multiple scientific questions that need to be answered from a clinical trial to establish the benefit of a treatment intervention. Consequently, there are two types of analyses that are important in understanding the impact of treatment intervention.

Sensitivity analyses are a series of analyses targeting the same estimand. This means there is usually no change to the treatment, population, variable, handling of intercurrent events, and population-level summary measure. These analyses assess different assumptions of the statistical methods to explore the robustness of the statistical inference, such as the main estimator and deviations from its underlying modeling assumptions or limitations in the data. Note that this redefines the term “sensitivity analyses” compared to its traditional pre-addendum use.

Supplementary analyses target a different estimand of interest. This section focuses on sensitivity and supplementary analysis for PFS and OS, illustrated by the case studies introduced in Section 2. Please note the recommendation below is linked to the GALLIUM study only. In general, whether analyses are sensitivity or supplemental, or appropriate at all, is dependent on the particular context and strategy. Different analyses may serve as the primary analyses for different stakeholders. A supplementary analysis for one stakeholder or in particular context may be the primary or a sensitivity analysis for another stakeholder or in another context.

4.1 Progression free survival

In this section, we provide the working group’s recommendations how to embed these analyses for PFS in the addendum’s framework of primary, sensitivity, and supplementary analyses. Progression free survival is defined as the time
from randomization until death from any cause or objective disease progression. PFS is an acceptable endpoint for oncology treatment approval in many tumor types.\textsuperscript{15} Table 3 summarizes common analyses performed for PFS using Gallium study for illustration purpose in addition to the main analysis as stated in Section 3. The primary analysis is a stratified Cox regression analysis based on investigators' assessed PFS irrespective of any subsequent anti-cancer therapy use. The main analysis is aligned with the scientific question presented in Table 2. The concept described in here applies also to related endpoint definitions like EFS in AML.

4.2 Overall survival

Now, we provide some recommendations how to embed addendum’s framework of primary, sensitivity, and supplementary analyses for OS. OS is defined as the time from randomization until death from any cause. Survival is considered the most reliable cancer endpoint, and when trials can be conducted to adequately assess survival, it is the preferred endpoint.\textsuperscript{15,16} OS is often performed based on survival data collected in the study for all patients. Table 4 lists common analyses performed for OS in addition to the main analysis as stated in Section 3, which is the stratified Cox regression analysis using the survival data collected in the entire study duration ignoring any subsequent anti-cancer therapy use.

5 IMPACT ON TRIAL DESIGN, DATA COLLECTION, AND DATA ANALYSIS

The design of a clinical trial needs to be aligned with the choice of estimand or estimands which reflects the primary objectives of the clinical trial. The agreed estimands dictate the data that needs to be collected. Different estimands might have different requirement of data collection. More importantly, a clinical trial may have multiple estimands, which implies that the data collection should be determined by the need to address them all (See Scharfstein,\textsuperscript{26} section A.4 of ICH E9 (R1)\textsuperscript{27}).

For Gallium study, the study objective was to evaluate the efficacy of G-chemo followed by G-maintenance therapy compared with R-chemo followed by R-maintenance therapy in patients with previously untreated advanced FL, as measured by investigator-assessed PFS. The primary estimand was the hazard ratio of G-chemo followed by G-maintenance therapy over R-chemo followed by R-maintenance therapy in the target population, regardless of the subsequent anti-lymphoma therapy use or whether patients discontinued planned treatment. The study design was aligned with the primary objective. If there is another objective to investigate if G-maintenance is more efficacious than R-maintenance, then at the end of induction phase, for those who had PR or better, they should be re-randomized into G-maintenance or R-maintenance. Similarly, for the 2nd and 3rd case study in patients with multiple myeloma and AML respectively, the study objective was to evaluate the efficacy of entire treatment sequence (induction followed by consolidation and then maintenance), the design described in Section 2 was the most efficient one. However, the design will not allow the causal inference of added effect of the new treatment in the maintenance. A potential challenge with this design is that the observed PFS benefit may be driven by the benefit of new drug in induction/consolidation only. A 2nd randomization after induction/consolidation would be needed if it is of interest to characterize the maintenance effect.

In both Gallium and RATIFY studies, the strategy for intercurrent events such as subsequent therapy and treatment discontinuation is as treatment policy. For the multiple myeloma case study, even though the primary estimand used hypothetical strategy for the intercurrent event of subsequent therapy, a supplementary analysis corresponding to treatment policy strategy for subsequent therapy was planned. Therefore, data for disease assessment should be collected until patients progress, or die if no progression or study end, whichever is earlier. The focus of discussion has been on the estimand with PFS/OS endpoints so far. In fact, PFS2, which is defined from randomization to progression corresponding to the first subsequent therapy or death, is also of interest from regulatory perspective in most of cases. Consequently, disease assessment, especially PD assessment on the next line subsequent therapy should also be collected.

From data analysis perspective, an analytic approach, or estimator, should be aligned with the given estimand and should be able to provide an estimate on which a reliable interpretation can be based. This implies that any assumptions should be explicitly stated, and sensitivity analyses should be carefully described to assess the robustness of the
| Analysis method                                              | Objective and key assumptions                                                                 | Sensitivity or supplementary analysis | Rationale | Recommendation |
|--------------------------------------------------------------|------------------------------------------------------------------------------------------------|--------------------------------------|------------|----------------|
| Stratified cox regression using PFS by independent review committee (IRC-PFS) | Impact of potential reader-evaluation bias in the treatment effect estimate. IRC-PFS is perceived as less biased compared with INV-PFS since this is an open-label study. As in the main analysis, patients are assumed to be consistently followed with disease assessment beyond subsequent therapy | Sensitivity analysis                  | Analysis corresponds to the same primary estimand as stated in the Table 2. It is considered as data limitation for disease assessment | Same disease assessment criteria should be used for both investigators’ assessment and IRC assessment. IRC assessment may introduce bias due to informative censoring. The bias due to informative censoring may be worse than the potential bias by investigators’ assessment in particular in blinded studies. |
| Unstratified Cox regression                                  | Impact of heterogeneity in placebo effect for meaningful strata on treatment effect estimate. The assumption refers to common baseline risk among different strata | Sensitivity analysis                  | Same estimand as in Table 2. The main analysis assumes different baseline risk across different strata. This analysis checks the heterogeneity assumed in the main analysis and explores how robust the overall findings are | Useful analysis for checking robustness of the main analysis |
| Same analysis as the main analysis except for imputing the event time of subjects considered to have problematic informative censoring | Impact of informative censoring on treatment effect. Complete and partial imputation of the informative drop-out (e.g., drop out due to adverse event, subject choice) using missing at random assumption | Sensitivity analysis                  | To demonstrate the impact of violations in necessary assumptions, and assesses robustness of the primary analysis result as calculated from imputed data as compared with un-imputed data | Checking if the distribution of censoring was approximately random or balanced between the arms |
| Censoring for 2 or more consecutive missing assessment(s)    | Impact of non-informative missing on treatment effect. Patients could be assessed as PD during the period of missing disease assessment | Sensitivity analysis                  | Targets the same estimand. The main analysis assumes patients who had missing disease assessments did not have PD before resuming disease assessment. The sensitivity analysis assumes all assessments after consecutive missing assessment period are no longer reliable because the patient could have PD already. The analysis treats missing assessments as administrative termination for disease assessment and check how robust the overall findings are to the assumption of no PD during the period of missing disease assessment | It is controversial. This could be considered as a supplementary analysis because we treat missing assessments as an intercurrent event and apply a hypothetical estimand strategy “of a world where these assessments would not have been missing.” This is especially true when there is a known event such as health care facility closure temporally due to pandemic. In this case, it corresponds to a different estimand (hypothetical) |
results to the underlying assumptions. In Gallium study, analyses performed on PFS included different population (FL only and overall population), different censoring schemes, and with or with adjustment for stratification and/or baseline covariate. Some of them were targeted on the same estimand and should be clarified as sensitivity analysis but others should be specified as supplementary analysis to avoid difficulty in interpretation. The subtlety between supplementary and sensitivity analysis is illustrated in Section 4. For the 2nd case study in patients with multiple myeloma, we may not be able to characterize the benefit magnitude of the new drug when added into the lenalidomide for the maintenance for the given study design, however, it is still of interest to demonstrate adding the new drug to lenalidomide during the maintenance does contribute to the overall benefit as the entire treatment sequence. Therefore, supplementary analyses such as treating maintenance as the time dependent covariate or new PFS with reference time as the start of maintenance might be to be included in the analysis plan. The treatment regimen in the 3rd case study in

| Analysis method | Objective and key assumptions | Sensitivity or supplementary | Rationale | Recommendation |
|-----------------|--------------------------------|-----------------------------|-----------|----------------|
| Analysis using per-protocol (PP) population | PP population represents a subset of patients, which adhere to the protocol defined condition. | Supplementary analysis | This analysis addresses the population attribute of estimand. PP population definition often depends on the outcome of treatment, such as completing a minimum of six cycles of treatment. The E9 R1 guidance concerns about the interpretability of per protocol set (PPS) analyses. It states ‘In respect of the framework presented in this addendum, it may not be possible to construct a relevant estimand to which analysis of the PPS is aligned’ | It is highly recommended not to have any such analysis. Patients in PP population do not represent random samples of the target population as defined by the eligibility criteria. Consequently, the interpretation is very difficult and such analysis is not useful. |
| Censoring at subsequent therapy | Subsequent therapy impacts the interpretation of treatment effect | Supplementary analysis | Different strategy for intercurrent event and hence it corresponds to different estimand. It is a hypothetical strategy and estimation via simple censoring assumes non-informative censoring at subsequent therapy use | Useful analyses to understand the treatment effect from different perspectives. In the past, the FDA could accept this as the main analysis while the EMA often preferred treatment policy strategy for the intercurrent event of subsequent therapy |
| Covariate-adjusted analysis (multivariate Cox regression analysis) | Impact of important baseline covariates on treatment effect. Covariates adjustment reduce variability of the effect estimate | Supplementary analysis | The analysis addresses a different scientific question. That is, what is the treatment benefit of Obinutuzumab in terms of PFS for patients who have covariates valued at population means. Covariate adjusted multivariate Cox regression provides a conditional treatment effect rather than marginal treatment effect. | Interpretation should be careful, and results may not be consistent since it does not address the same estimand |
| Analysis method | Objectives and key assumptions | Sensitivity or supplementary | Rationale | Recommendation |
|-----------------|--------------------------------|-----------------------------|-----------|----------------|
| Censor at subsequent anti-cancer therapy | Patients who received subsequent anti-cancer therapy have the same risk as those who did not receive subsequent anti-cancer therapy | Supplementary analysis | Using the hypothetical strategy to account for the intercurrent event, new anticancer therapy, which is a new estimand. | The underlying clinical question of this analysis may not be relevant from regulatory perspective because the analysis addresses the treatment difference under assumption that no subsequent therapy is available. It may be of interest from market access or reimbursement perspective. |
| Rank preserving structural failure time (RPSFT) model | Treatment effect is the same regardless of when the experimental treatment is initiated | Supplementary analysis | The analysis is to address a different question: what the treatment benefit is if patients in the Rituximab arm are not allowed to receive Obinutuzumab after PD. The treatment effect is expressed as an acceleration factor through G-estimation or by estimating hazard ratios using reconstructed data through the Cox model. Different strategy (an alternative hypothetical strategy) for intercurrent of treatment crossover (control to experiment arm). Hence it corresponds to different estimand. | Useful analysis to adjust for crossover. This method is advantageous as it preserves randomization, does not require information on covariates and can handle larger proportion of patient switching. This analysis is often of interest to the HTA for reimbursement decision |
| Inverse probability censoring weighting (IPCW) analysis | All common predictors for survival and/or the probability of treatment switching are measured | Supplementary analysis | The corresponding scientific question is “what the treatment difference would be if subsequent anti-cancer therapies are not available?” It is one of hypothetical strategies used for intercurrent event of subsequent anti-cancer therapy. | Clinical interpretation is very different from the main OS analysis. However, it provides some insight in understanding the impact of treatment intervention. This analysis is often of interest to the HTA for reimbursement decision |
| Two-stage approach\textsuperscript{23,24} | OS benefit adjusted for treatment crossover at a specific disease-related time-point progression (PD). It requires not only the complete data collection at PD, but also the assumption that switching occurs soon after PD | Supplementary analysis | The treatment effect for those patients who cross over is removed, that is, survival time after progression is adjusted using a hypothetical strategy. The treatment effects are estimated separately for crossover patients from control group and for patients originally randomized to the investigational drug group using an accelerated failure time (AFT) model | Like other counterfactual hypothetical strategies (RPSFT, IPCW) above, the two-stage approach assumes no unmeasured confounders.\textsuperscript{25} This assumption can be evaluated by running a full and reduced model for comparison. |
patients with AML is even more complicated. Like the other 2 trials, it has multiple treatment phases. In addition, transplantation was performed at the discretion of the investigator. The contribution of transplantation to the treatment strategy may be of interest. Supplementary analysis to tease out the effect of transplantation should be included in the analysis plan even if transplantation itself is part of the treatment regimen.

6 | CONCLUSION AND DISCUSSION

The estimand framework is an efficient tool to ensure consistency between the scientific question and the definition of the study objectives. Per E9(R1), intercurrent events are not to be thought of as a drawback to be avoided. Discontinuation of treatment, changing treatment, etc. are part of clinical practice and are part of clinical trials. The framework ensures transparency in unfavorable yet unavoidable situations in clinical trials (“intercurrent events”). The outcome of studies that are following the estimand framework can be interpreted in a consistent manner. We highly recommend estimand(s) to be included in the study protocol and details including sensitivity analyses and supplementary analyses to be included in the statistical analysis plan (SAP) for the study. Some sensitivity analysis approaches such as tipping point analysis recommended in the estimand guidance document but not elaborated in this paper may also be considered in the SAP. Recommended template language can be found in Data S1.

The three case studies discussed in this paper represent typical treatment regimens and study designs for drug development in hematologic oncology. They share a common feature of having multiple treatment phases and the treatment in the follow-up phase may depend on the treatment effect on the previous phase. One potential issue with multiphase treatment is the violation of proportional hazard assumption for PFS analysis. Assumption of proportional hazard should be checked whenever Cox regression model is used. For non-proportional HR, the traditional Cox regression model to characterize the treatment benefit using HR may be difficult to interpret. Other analysis methods with different population level summary such as restricted mean survival time (RMST) analysis and responder analysis using pre-specified milestone PFS/OS proportion may need to be adapted. Common intercurrent events and potential strategies are discussed in detail for each case study. But the choice of strategies should depend on the disease as well as the mechanism of action of the drug. We only illustrated one estimand corresponding to the primary objective in each case study. However, one estimand may not be enough for the study objective. For example, in the 2nd case study, characterizing the benefit of the new drug when adding in the standard care regimen in induction phase, induction/consolidation phase are important together with the overall benefit for the entire treatment sequence with long term follow-up are equally important. Especially from regulatory perspective, the observed overall treatment benefit is not driven by one single phase. The formulation of the scientific question of interest requires an a priori discussion among stakeholders to tailor the estimand definition and the treatment strategy attribute accordingly: To assess the contribution of each treatment phase of a sequential treatment strategy might require re-randomization of eligible patients when starting a subsequent treatment phase. This might operationally be not feasible in many hematologic malignancies with rare patient populations.

Another challenge introduced in the Gallium study is the multiplicity control. It is not uncommon to include multiple tumor types of the same pathology in one study (both FL and MZL in Gallium study). Analysis on different population corresponds to different estimands. So, the significance level for the analysis associated with one estimand must be considered in the context of the analysis associated with other estimands.
In essence, the estimand framework is about an estimation (what to estimate and how to estimate). Hypothesis testing is also important for decision making, and the same analytical approach may be used for both hypothesis testing and estimation of treatment effect.

Recently, CAR-T therapies have shown promising results for multiple hematological cancer types. The development of CAR-T therapies poses unique challenges from statistical analysis perspective. For example, bridging treatment is usually allowed during the manufacturing period and conditioning lymphodepletion chemotherapy is required before infusion of CAR-T cells. For single arm study, should the response rate be provided for all patients who received CAR-T infusion or for all patients who had apheresis? Similarly, what is the appropriate way for measuring treatment benefit for randomized control trials. There will not be any ambiguity for the treatment effect associated with the defined “treatment” (strategy a: sequence of bridging treatment, conditioning therapy, and infusion of CAR-T versus strategy b: infusion of CAR-T) under estimand framework. Emphasis of this paper is placed on the recommendation of description of estimands and careful selection of sensitivity analyses and supplementary analyses for hematological trials. Data collection and analysis should also be aligned in coherent manner to avoid disconnect between trial objectives and estimands. The template language on estimands for protocols and SAPs in Data S1 can serve as detailed instructions/recommendations.

ACKNOWLEDGEMENTS
We would like to thank our cross-industry working group member Marie-Laure Casadebaig from Celgene, Viktoria Stalbovskaya from Merus, and Evgeny Degtyarev from Novartis for their inputs and comments.

CONFLICT OF INTEREST
The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT
Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID
Steven Sun https://orcid.org/0000-0003-4603-2043
Kaspar Rufibach https://orcid.org/0000-0002-2634-1167
Satrajit Roychoudhury https://orcid.org/0000-0003-4001-3036

REFERENCES
1. Faulkner KG, Yamamoto R, Conrad E. Estimand framework: what it is and why you need it. Appl Clin Trials. 2020. https://www.appliedclincialtrialsonline.com/view/estimand-framework-what-it-and-why-you-need-it.
2. Palmisiano N, Harverkos B, Gaballa S, Filicko-O’Hara J, Porcu P, Kasner M. Clinical trials in hematologic malignancies. Oncology Clinical Trials: Successful Design, Conduct, and Analysis. New York: demosMREDICAL, Springer Publishing; 2018.
3. Roussel M, Hebraud B, Hulin C, et al. Twice weekly induction with ixazomib-lenalidomide-dexamethasone (IRD) combination followed by extended IRD consolidation and lenalidomide maintenance in transplant eligible patients with newly diagnosed multiple myeloma (NMM): a phase 2 study from the Intergroupe Francophone Du Myelome (IFM 2014-03). Blood. 2019;134:3159.
4. Salwender H, Bertsch U, Weisel K, et al. Rationale and design of the German speaking myeloma multicenter group (GMMG) trial HD6: a randomized phase III trial on the effect of elotuzumab in VRD induction/consolidation and lenalidomide maintenance in patients with newly diagnosed myeloma. BMC Cancer. 2019;19:504.
5. Manitz J, Kan-Dobrosky N, Buchner H, et al. Estimands in clinical trials with treatment switching; 2020. Under revision with submission to Pharm Stat.
6. Ellis LM, Bernstein DS, Voest EE, et al. American Society of Clinical Oncology perspective: raising the bar for clinical trials by defining clinically meaningful outcomes. J Clin Oncol. 2014;32:1277-1280.
7. Johnson JR, Williams G, Pazdur R. End points and United States Food and Drug Administration approval of oncology drugs. J Clin Oncol. 2003;21:1404-1411.
8. Smith BD, DeZern AE, Bastian AW, Durie BG. Meaningful endpoints for therapies approved for hematologic malignancies. Cancer. 2017;123:1689-1694.
9. Zhang J, Grubor V, Love CL, et al. Genetic heterogeneity of diffuse large B-cell lymphoma. Proc Natl Acad Sci U S A. 2013;110(4):1398-1403.
10. Marcus R, Davies A, Ando K, et al. Obinutuzumab for the first-line treatment of follicular lymphoma. N Engl J Med. 2017;377:1331-1344.
11. Stone R, Mandrekar S, Sanford B, et al. Midostaurin plus chemotherapy for acute myeloid leukemia with a FLT3 mutation. N Engl J Med. 2017;377:454-464.
12. Holzhauer B, Akacha M, Bermann G. Choice of estimand and analysis methods in diabetes trials with rescue medication. Pharm Stat. 2015;14(6):433-447.
13. Akacha M, Bretza F, Rubergb S. Estimands in clinical trials – broadening the perspective. Stat Med. 2017;36:5-19.
14. Mehrotra DV, Hemmings RJ, Russel-Cohen E. Seeking harmony: estimands and sensitivity analyses for confirmatory clinical trials. Clin Trials. 2016;13(4):456-458.
15. FDA guidance. Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics Guidance for Industry. 2018. Accessed December 2018. https://www.fda.gov/media/71195/download.
16. EMA/CHMP guidance. Guideline on the evaluation of anticancer medicinal products in man. 2017. Accessed January 2019. https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-evaluation-anticancer-medicinal-products-man-revision-5_en.pdf.
17. Fleischcr F, Gaschler-Markefski B, Bluhmki E. How is retrospective independent review influenced by investigator-introduced informative censoring: a quantitative approach. Stat Med. 2011;30(29):3373-3386.
18. Amit O, Mannino F, Stone A, et al. Blinded independent central review of progression in cancer clinical trials: results from a meta-analysis. Eur J Cancer. 2011;47(12):1772-1778. https://doi.org/10.1016/j.ejca.2011.02.013.
19. Zhang JJ, Chen H, He K, et al. Evaluation of blinded independent central review of tumor progression in oncology clinical trials: a meta-analysis. Ther Innov Regul Sci. 2013;47:167-174.
20. Liu Y. Sensitivity analyses for informative censoring in survival data: a trial example. J Biopharm Stat. 2017;27(4):595-610.
21. Degtyarev E, Rufibach K, Shantu Y, et al. Assessing the impact of COVID-19 on the clinical trial objective and analysis of oncology clinical trials—Application of the estimand framework. Stat Biopharm Res. 2020;12:43-53.
22. Latimer N, Abrams K, Lambert P, et al. Adjusting survival time estimates to account for treatment switching in randomized controlled trials—an economic evaluation context: methods, limitations, and recommendations. Med Decis Making. 2014;34(3):387-402.
23. Alshurafa M, Briel M, Akl EA, et al. Inconsistent definitions for intention-to-treat in relation to missing outcome data: systematic review of the methods literature. PLoS One. 2012;7(11):e49163. https://doi.org/10.1371/journal.pone.0049163.
24. Watkins C, Huang X, icholas LN, Tang Y, Wright E. Adjusting overall survival for treatment switches: commonly used methods and practical application. Pharm Stat. 2013;12(6):348-357.
25. Scharfstein DO. A constructive critique of the draft Ich E9 Addendum. Clin Trials (London, England). 2019;16(4):375-380. https://doi.org/10.1177/1740774519853566.
26. ICH E9 Working Group. ICH E9 (R1): Addendum on Estimands and Sensitivity Analysis in Clinical Trials to the Guideline on Statistical Principles for Clinical Trials. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use; August 2019. https://database.ich.org/sites/default/files/E9_Guideline.pdf.
27. Rufibach K. Treatment effect quantification for time-to-event endpoints–Estimands, analysis strategies, and beyond. Pharm Stat. 2019;18(2):145-165.
28. Bengoudifa B, Weber H, Gathmann I, Hoenekopp A, Berkowitz N. Estimand framework in AML based on a randomized, placebo-controlled, phase 3 study investigating the effect of adding midostaurin to standard chemotherapy in patients with newly diagnosed AML without a FLT3 mutation. Blood. 2018;132:5196.
29. Byod A, Kittelson J, Gillen D. Estimation of treatment effect under nonproportional hazards and conditionally independent censoring. Stat Med. 2012;31(28):3504-3515. https://doi.org/10.1002/sim.5440.
30. Gregson J, Sharples L, Stone GW, Burman CF, Óhnm F, Pocock S. Nonproportional hazards for time-to-event outcomes in clinical trials: JACC review topic of the week. J Am Coll Cardiol. 2019;74(16):2102-2112.
31. Leuchs A-K, Zinserling J, Brandt A, Wirtz D, Benda N. Choosing appropriate estimands in clinical trials. Ther Innov Regul Sci. 2015;49(4):584-592.
32. Leuchs AK, Brandt A, Zinserling J, Benda N. Disentangling estimands and the intention-to-treat principle. Pharm Stat. 2017;16(1):12-19.
33. Degtyarev E, Zhang Y, Sen K, et al. Estimands and the patient journey: addressing the right question in oncology clinical trials. JCO Precis Oncol. 2019;3:1-10.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Sun S, Weber H-J, Butler E, Rufibach K, Roychoudhury S. Estimands in hematologic oncology trials. Pharmaceutical Statistics. 2021;20:793–805. https://doi.org/10.1002/pst.2108