



Clinical Trial Drug Safety Assessment for Studies and Submissions Impacted by COVID-19

Mary Nilsson^a, Brenda Crowe^a, Greg Anglin^b, Greg Ball^c, Melvin Munsaka^d, Seta Shahin^e, and Wei Wang^b

^aEli Lilly and Company, Indianapolis, IN; ^bEli Lilly Canada Inc., Toronto, ON, Canada; ^cMerck & Co., Inc., Rahway, NJ; ^dAbbVie Inc., North Chicago, IL; ^eAmgen Inc., Thousand Oaks, CA

ABSTRACT

In this article, we provide guidance on how safety analyses and reporting of clinical trial safety data may need to be modified, given potential impact from the COVID-19 pandemic. Impact could include missed visits, alternative methods for assessments (such as virtual visits), alternative locations for assessments (such as local labs), and study drug interruptions. Starting from the safety analyses typically included in Clinical Study Reports for Phase 2–4 clinical trials and integrated submission documents, we assess what modifications might be needed. If the impact from COVID-19 affects treatment arms equally, analyses of adverse events from controlled data can, to a large extent, remain unchanged. However, interpretation of summaries from uncontrolled data (summaries that include open-label extension data) will require even more caution than usual. Special consideration will be needed for safety topics of interest, especially events expected to have a higher incidence due to a COVID-19 infection or due to quarantine or travel restrictions (e.g., depression). Analyses of laboratory measurements may need to be modified to account for the combination of measurements from local and central laboratories.

ARTICLE HISTORY

Received April 2020
Accepted July 2020

KEYWORDS

Clinical trial; Coronavirus; COVID-19; Harms; Safety analysis; Safety review

1. Introduction

The pandemic of coronavirus disease (COVID-19) has had broad impact on ongoing clinical trials. Guidance has been released by various stakeholders and regulatory agencies, for example, Association of Clinical Research Organizations (2020), European Medicines Agency (2020a, 2020b), McDermott and Newman (2020), U.S. Food and Drug Administration (2020), Meyer et al. (2020), and Akacha et al. (2020) to address some of the challenges. As indicated by these guidance documents and publications, challenges may arise from quarantines, site closures, travel limitations, or other considerations if site personnel or trial subjects become infected with COVID-19. These challenges may lead to difficulties in meeting protocol-specified procedures, including administering or using the investigational product or adhering to protocol-mandated visits and laboratory/diagnostic testing. Thus, study drug interruptions could be more common and longer in duration, and missed visits and patient discontinuations could be more common. For most, but not all, safety outcomes, alternative methods for safety assessment could be implemented, for example, phone contact, virtual visits, alternative locations for assessment (including local labs or imaging centers), that lead to differences in how patient information is received and recorded. The method of obtaining information should be considered carefully as there may be limitations in interpretation depending on the collection approach (PHUSE 2017).

To assess how safety analyses and reporting of clinical trial safety data may need to be modified, we started with recommended safety analyses proposed in white papers and a workshop, created as part of an FDA/PHUSE collaboration

(PHUSE 2013, 2015, 2017, 2019), with a focus on Phase 2–4 ongoing clinical trials. PHUSE is an independent, not-for-profit organization run by volunteers, established in 2004. The organization provides a global platform for discussion of topics encompassing the work of data managers, biostatisticians, statistical programmers, and information technology (IT) professionals. In 2012, PHUSE and the FDA created a collaboration to provide an open, transparent and collaborative forum in a noncompetitive environment in which academics, regulators, sponsors, technology providers, and other stakeholders could address computational science needs in support of medicinal product development and regulatory review (Rosario et al. 2012). As part of this collaboration, a project team was formed to create white papers containing recommended safety analyses for industry standardization and education. These white papers have gone through a robust, public peer-review process. More information on PHUSE and the collaboration can be found at phuse.eu. We have chosen recommendations from the PHUSE white papers and workshop as our starting point, as they likely reflect the types of analyses that would be planned for ongoing studies that are part of a clinical development program. Whether or not product teams have implemented plans in accordance with PHUSE recommendations (and even if they are not familiar with them), the principles still apply.

In this article, using standard safety analyses as a framework, we examine potential impact from COVID-19 on the scientific evaluation of safety data from clinical trials overlapping in time and geography with the pandemic. Guidance is provided on how to simply and properly reframe the analyses. We recognize that when assessing the potential impact of COVID-19 on safety

analyses, discussion could evolve to consider updating safety planning to use the estimand framework (Unkel et al. 2019) if not already incorporated, and/or to use alternative methods proposed in recent literature (e.g., Stegherr et al. 2019 and references therein). While we acknowledge the need for future research on methods for safety analyses and implementation of the estimand framework for safety in clinical trials, to assist cross-functional study teams with the most immediate needs, we have chosen to focus on the safety analyses that are likely already planned and the associated purpose without introducing estimand-related language. We focus on analyzing the observed data. As noted by Hemmings (2020), with respect to missing safety data, making predictions for toxicities from a statistical model seems difficult and conceptually undesirable.

When thinking through the planned safety analyses, and potential additions or changes that might be needed due to the impact of COVID-19, it is important to consider the purpose of the analyses. While safety analyses have many purposes, the two main purposes are to help determine which adverse events are causally related to the drug (i.e., are adverse drug reactions [ADRs]) and to quantify the risk of ADRs in labeling or in other risk communications. This article intends to address these two main purposes. We understand that these extraordinary circumstances might provoke additional interest beyond these two main purposes. For example, are there differences in event reporting by patients in virtual visits versus live visits? And what are risk factors associated with COVID-19 infection? Since these types of analyses are likely better addressed as separate reports combing data from multiple studies across compounds, these are out-of-scope. While this article has been written by statisticians, we acknowledge medical assessment of the analyses, as well as medical safety review of individual cases, with understanding of medication class effects, biological plausibility, and other clinical considerations, are also important (CIOMS Working Groups III and V 1999; CIOMS Working Group VI 2005; PHUSE 2019).

We start with a brief discussion on the collection of data to assess the impact of COVID-19 on study conduct, followed by a discussion on assessing the impact. The majority of the article then focuses on the most common analyses of clinical safety data, organized to correspond to sections in a Clinical Study Report or regulatory submission. We end with a discussion on the impact on labeling and an overall summary. While the details of the disruption of clinical trials by COVID-19 are expected to be so diverse that no single solution will be appropriate for all trials (Collins and Levenson 2020), we hope the suggestions provided in this article will facilitate discussion within study and product teams.

2. Data Collection and Assessing COVID-19 Impact

2.1. Data Collection Assumptions

This article is not intended to provide details on modifications to data collection that might be needed in ongoing studies impacted by COVID-19. Other groups, such as the Clinical Data Interchange Standards Consortium (CDISC), are addressing this (CDISC 2020). Given recent guidance from the European Medicines Agency (2020a, 2020b) and the U.S. Food and Drug

Administration (2020), it seems reasonable to expect that studies will have data collected to assess the impact of COVID-19 on study conduct. The data may be captured through existing case report forms through instructions to sites on where and how to convey COVID-19 impact information (e.g., specify fields) and/or from new case report forms developed specifically to capture COVID-19 impact information (Akacha et al. 2020; Meyer et al. 2020). For safety analysis purposes, we make the following assumptions:

- There will be a way to identify patients who have visits impacted by COVID-19. This information may be needed for some sensitivity analyses and/or for safety topics of interest.
- If patients cannot attend visits—due to quarantines, site closures, and travel limitations—key safety data collection (including adverse events [AEs], serious adverse events [SAEs], and critical labs) will continue through alternative means (such as through phone contacts, virtual visits, and local labs). In the case that a longer-than-usual time elapsed since safety data were collected, patients may remember fewer AEs, but should remember the most impactful events.

2.2. Assessing COVID-19 Impact

As noted earlier, multiple sources have delineated the potential impacts of COVID-19 upon clinical trial data. Generally speaking, the biggest impacts and safety evaluation challenges are due to quarantines, site closures, and stay-in-place orders leading to additional discontinuations, missing data due to missed visits, treatment interruptions, procedures performed differently to enable remote assessments, and safety monitoring challenges. TransCelerate Biopharma Inc. (2020) created a document that offers points to consider and guiding principles in how to describe the impact in Clinical Study Reports. Generally, the description will include whether the study or sites were stopped or paused, whether dosing was paused, whether visits were conducted in alternative locations, whether any vendors were added or changed, what assessments were reported by phone versus in person. We agree that the impact should be fully described in a relatively self-contained way (such as the Summary of Changes section or an appendix), as to not obscure or detract from the results and conclusion of the study, unless the study stopped or failed as a result of the pandemic.

To assess the safety of study participants, a fundamental concern is whether there are any variables that might apply differentially across treatment groups, in a way that could influence the conclusions. Such variables include patient characteristics (such as gender, age, race) but also include aspects of study conduct (such as discontinuations, missed visits, protocol deviations, and number of missed doses). For example, discontinuations should be examined, and not only with regard to proportions of discontinuations in each treatment group, but also with respect to any patterns in timing of discontinuations and follow-up time. The potential impact of any differences should be considered.

In the context of characterizing impacts from COVID-19, evaluation of these patient characteristics and other aspects are still relevant and appropriate. However, additional evaluations may be warranted, such as assessing the proportion of visits

performed remotely rather than in person, to decide if there are meaningful differences across treatment groups.

Another aspect of assessing the impact of COVID-19 pertains to power. The impact of COVID-19 could lead to fewer events associated with investigational product reported than what otherwise might have been obtained, with a related reduction in power to detect true treatment effects. As highlighted by several authors (Foulkes 2007; Singh and Loke 2012; Nautiyal, Rastogi, and Gamperl 2015), individual clinical trials are typically designed based on efficacy outcomes resulting in limited power for safety outcomes. However, it is possible that teams have considered power for specific safety outcomes (Foulkes 2007; Crowe et al. 2009), or that the study itself may have a safety outcome as its primary endpoint. In these cases, the power calculations will need to be reassessed. It is important that any reassessment of power and sample size be carefully thought through, taking into account the timing of COVID-19 on the stage of the study in terms of accrual and follow-up along with other important considerations. These considerations may include adaptation, blinded or unblinded assessment, treatment-effect homogeneity assessment, early readout potential, and use of Bayesian methods. Kunz et al. (2020) provide a detailed discussion of these considerations and methods for resizing trials along with an R Shiny tool that implements the methods in the context of COVID-19.

3. Analyses of Adverse Events and Concomitant Medications

3.1. Comparing Adverse Events Between Treatments

For fixed-duration studies with a similar distribution of follow-up times among treatment groups, comparing percentages of patients with specific treatment-emergent adverse events (TEAEs), SAEs, and AEs leading to study drug discontinuation among treatment arms (such as for investigational arms vs. placebo) is generally useful and commonly planned. The intent of these analyses is to determine if there are any important imbalances between treatment groups disfavoring study drug, that could be suggestive of a causal relationship (Crowe et al. 2013; Ma et al. 2015; PHUSE 2019). If substantially more study patients were to discontinue study or study treatment during the controlled period in one treatment arm versus another, then different analytical approaches would be needed. See Section 10.9 of PHUSE (2017) and Stegherr et al. (2019) (and the references therein) for a discussion of some pitfalls of percentage-based methods and possible alternative methods. With the COVID-19 pandemic, it can be expected that more patients will discontinue early or have periods in which study drug has been interrupted. Unless the impact is considerably different across treatment arms, analytical plans can generally remain unchanged for purposes of detecting imbalances between treatment groups. However, if the impact of COVID-19 is large, percentages from all patients may not be useful for quantification of risk purposes (see Section 8).

3.2. Summarizing Adverse Event Data Without a Control

Within statistical analysis plans for extension studies or studies with an extension period, there are often plans to summarize

counts and percentages and/or exposure-adjusted incidence rates (EAIRs) for adverse events for the investigational product. These studies and study periods often lack a control arm. Generally, the intent of these summaries is to provide an easy way to identify some of the rarer events that might require case review. These summaries are not usually used for any comparisons. If they were used for comparisons, caution would be required. This would be true even before introducing any issues arising from the COVID-19 pandemic, though COVID-19 has amplified the issues associated with comparing uncontrolled data to other sources. For example, with the COVID-19 pandemic, there could be a substantial amount of time in which patients were off investigational product (e.g., during a quarantine). Percentages and EAIRs could therefore be under-estimated. As another example, percentages and EAIRs could be impacted due to differences in ascertainment of adverse events (e.g., a phone call instead of a site visit). Additionally, events associated with COVID-19 (e.g., fever, cough) or events associated with physical/social isolation or economic hardships (e.g., depression) might appear at a more frequent rate. Consequently, comparing an EAIR from studies/integrated summaries impacted by COVID-19 with an EAIR from the literature or other source could be even more problematic than usual. While potential overestimation and underestimation from COVID-19 impact applies to all data, the issue is particularly pertinent when interpreting uncontrolled data against literature or other external sources. When it is necessary to compare an EAIR with another source to use as a background rate, summarizing up to COVID-19 impact or by COVID-19 subgroups (such as patients without impact and patients with impact) might be helpful. Similarly, as noted in Section 3.1, when data from studies impacted by COVID-19 is used for risk quantification, additional considerations are required (see Section 8).

3.3. Concomitant Medications

A summary of concomitant medication use by treatment arm is commonly planned. The primary purpose of this summary is to assess whether there is an imbalance in concomitant medications use among treatment groups that would be important to consider when reviewing adverse event summaries. With the COVID-19 pandemic, it is quite likely that there would be changes in the concomitant medication usage. However, since the main purpose of this summary is to detect imbalances among treatment arms, the planned summary and associated purpose does not need to change. An accurate estimate (e.g., for a product label) of concomitant medication use is rarely needed. Note, medication data would continue to play an important role in case reviews. See Section 6 for further information.

4. Analyses of Laboratory Data

When a central lab is normally used for a study, but local labs are subsequently used due to COVID-19, this would be an important impact of COVID-19 to consider. This situation can occur if patients are unable to attend a site visit (e.g., the site is at a hospital that is closed to clinical trial activities) but are able

to go to a different location for select laboratory measurements needed for safety monitoring.

If a local lab was sometimes used but the measurements were not brought into the study database, then analyses using central lab data would be conducted with less complete data than what would have otherwise been available. Since analyses using central lab data will be incomplete, interpretation will need to include a review of adverse event data for laboratory-related findings.

If a local lab is used and measurements are brought into the study database, then analytical plans for central tendency analyses will need to be updated to provide clarity on if and when local lab measurements and central lab measurements will be combined. For many lab analytes, combining local and central labs could help fill in the gaps, providing more complete data. However, for some analytes, directly combining the data may not be appropriate. Note, even when combining local and central labs, additional variability and uncertainty can be added into the data. See Section 4.2 for more details.

4.1. Comparing Percentages of Shifts to High/Low Between Treatments

For fixed-duration studies with a similar distribution of follow-up times among treatment groups, comparing percentages of patients shifting from normal/low to high and normal/high to low (sometimes referred to as treatment-emergent highs and lows) among treatment arms is commonly planned. Similar to comparing percentages for adverse event data, the intent of these analyses is to assess the imbalance among treatment arms (Section 3.1). As with adverse events, analytical plans can mostly remain unchanged unless the impact from COVID-19 is different across treatment groups. As previously noted, if the impact is different across treatment groups, different or additional analyses are likely warranted. For these summaries, combining measurements from local labs and central labs is generally appropriate, as long as the limits from the associated lab are used.

4.2. Simple Summary Statistics by Visit

Summarizing changes over time by treatment using simple statistics (e.g., boxplots by visit with means or mean changes below the plot) is commonly planned for individual studies. Using simple summary statistics could be problematic if data collection is impacted by COVID-19. During the pandemic there could be a substantial number of missed visits. Under these circumstances, reporting means based on a mixed model for repeated measures (MMRM) instead of simple means may be more appropriate. Moreover, if a local laboratory is used and the measurements are brought into the study database, the study team will need to decide which laboratory measurements can be combined. Alternatively, study teams can choose a different analytical approach that allows for combining laboratory measurements from different laboratories. A frequently used technique is to report the data as a percent above/below normal limits. Alternatively, a normalization method could be used to combine local and central labs in the analysis. See Section

6.2.4 from PHUSE (2013) and Karvanen (2003) for further information.

4.3. Comparing Changes to Minimum/Maximum Values Between Treatments

Comparing change to a minimum/maximum value among treatments is sometimes planned for individual studies and/or integrated summaries. As with comparing percentages among treatments, comparing changes to minimum/maximum values among treatments should be appropriate, unless the average number of measurements is very different among treatment arms. If local labs are used and the measurements are brought into the study database, the same considerations described in Section 4.2 apply for these analyses.

4.4. Hepatotoxicity

Typically, in submissions, there is an expectation to assess the potential for drug-induced hepatic injury (FDA 2009). As part of this evaluation, a plot of alanine aminotransferase (ALT) versus total bilirubin is often created (Senior 2014). The upper right quadrant ($>3 \times$ ULN ALT, $>2 \times$ ULN total bilirubin) is often referred to as “Hy’s law range” or “potential Hy’s law cases.” Identification of true Hy’s law cases requires additional considerations, but this plot can be used to graphically show whether a study drug has the potential to cause hepatic injury. For this assessment, every occasion of having an ALT $>3 \times$ ULN and total bilirubin $>2 \times$ ULN matters. If local labs are used and data are not brought into the study database, there is a potential for missing patients that would otherwise have been in the Hy’s law quadrant. Careful review of adverse event data is required. Certainly, it would be better if all the results for hepatic enzymes were brought into the study database. If local labs are used and data are brought into the study database, the limits from the local laboratory should be used to determine the multiple of the upper limit of normal.

5. Intrinsic Factors

For large individual studies and integrated summaries of safety, there are often plans to summarize percentages of common TEAEs by subgroups. These subgroups usually include gender, age categories, and race (see Figure 12.2 of PHUSE (2017)). Additional subgroups may be included, depending on the indication under study. Summarizing by COVID-19 subgroups (based on some objective measure of COVID-19 impact) will generally be more useful for specific safety topics of interest, than for general TEAEs (see Section 7). However, summarizing percentages of common TEAEs by COVID-19 subgroups should be considered if the COVID-19 impact considerably differs across treatment arms.

6. Case Reviews

In addition to assessing numeric imbalances between drug and control, ADR determination involves other factors, including medical review of individual cases. Case reviews are conducted

to assess the potential relationship with investigational product versus a concomitant medication versus other conditions the patient may have been experiencing. It is common to create individual patient displays (such as narratives and graphical patient profiles) to facilitate this review. These displays usually include demographics, study drug exposure, concomitant medications, AEs, labs, vital signs, and—when applicable—ECGs. For exposure, it is common to show start and stop dates of study drug. If study drug had been interrupted due to a COVID-19 quarantine or other reason, this should be reflected. If a visit had been impacted by COVID-19 in any manner, this should be reflected. Knowing dates for study drug exposure and knowing whether visits had been impacted in any manner would be helpful for these case reviews. For SAEs in particular, gathering and considering all available information is crucial for meaningful case reviews.

7. Safety Topics of Interest

While analytical planning for general safety assessment can largely remain the same (with some exceptions), special consideration is needed for safety topics of interest, particularly those that could have a higher incidence due to COVID-19 infection or due to the physical or social isolation caused by mandates to stay at home (e.g., depression). The cross-disciplinary team should discuss the possibility for additional or alternative methods that might be warranted. For example, summaries up to COVID-19 impact or by COVID-19 subgroups for some safety topics of interest would likely be warranted. Additionally, more complex methodologies (such as Kaplan–Meier plots, Cox proportional hazards models, and/or competing risk models) may need to be implemented. The need for additional methods will depend on the safety topic of interest and the extent of the COVID-19 impact (and impact from other factors). In choosing among alternative methods, it is important to try to connect the method with the eventual interpretation, and to understand the pros and cons of the various choices. A full discourse on these methods is beyond the scope of this article. Nevertheless, we offer some insights on one particular methodology, namely, competing risks. COVID-19 logistical problems are unlikely to introduce a need for competing risk analysis for a study for which no competing risk analysis was needed prior to the COVID-19 pandemic, unless a study has many deaths from COVID-19 infection. Competing risks are events that preclude or greatly alter the occurrence of the main event of interest. For example, if the event of interest is myocardial infarction, death from other causes would be considered a competing risk. Competing risks are different from other concurrent events in that they actually preclude the event of interest from happening, whereas events like early study discontinuation prevent the event from being *observed*. Various authors, for example, Allignol, Beyersmann, and Schmoor (2016), Bender, Beckmann, and Lange (2016), Geskus (2016), Hengelbrock et al. (2016), Proctor and Schumacher (2016), and Unkel et al. (2019) have written about the need to consider competing risks in the assessment of the risk/probability of adverse events. As noted by Allison (2018), standard Kaplan–Meier or Cox proportional hazards methods perform better for determining whether or not the

drug is causally related to the AE than methods that take into account the competing risk. Alternatively, if interest is in getting an accurate percentage of patients with the event (as for the product label), estimation methods that take competing risks into account may be useful.

For many safety topics of interest, there often plans to conduct analyses to further understand and characterize. These often include an assessment of event duration or whether a change in labs/vitals is transient versus persistent. With the COVID-19 pandemic, it is possible that less information would be available to assess these patterns in the data due to missing visits, early study discontinuation, early study drug discontinuation, or study drug interruption. Using group means to assess transient (vs. persistent) patterns always has the potential to be misleading (PHUSE 2019), but the potential could be even greater if discontinuations or interruptions were more common. Thus, for these assessments, using a display that graphically displays individual patient data is recommended. For events, see Appendix B in PHUSE (2017) for an example of a plot showing events over time (onset and duration). For labs/vitals, a spaghetti plot (plot of values [vertical axis] vs. time [horizontal axis] and connecting the dots chronologically with lines for each patient) can serve this purpose. In a spaghetti plot, symbols or color can be used for when a patient is on or off drug. While graphical displays of individual patient data cannot generally be included in labeling, they can be used as a source for general statements of characterization.

For some compounds, such as those used to treat autoimmune conditions, infections might already be a safety topic of interest. For these compounds and possibly others, considering the COVID-19 infection itself as a safety topic of interest could be warranted. This could arise if there is biological plausibility for a greater risk for infection and/or if there is imbalance in COVID-19 infections between treatment arms. When the COVID-19 infection itself is considered a safety topic of interest, a separate summary enumerating all patients with possible/probable/confirmed COVID-19 infections would be useful. If there are a substantial number of COVID-19 infections, additional summaries by subgroups (such as age, gender, and race) could also be useful, as COVID-19 affects some groups more than others. Of course, this is easier said than done. Fortunately, starting with MedDRA 23.0, COVID-19 preferred terms have been made available. Regardless, without universal testing of participants, identifying cases of COVID-19 infection is challenging. MedDRA MSSO is working on a Standardized MedDRA Query (MedDRA 2020), which will be very helpful. Cross-industry and regulatory collaboration on this topic, with appropriate medical and statistical representation, would be helpful to facilitate consistent analysis and reporting.

8. Quantification of Risk and Product Labeling

When communicating about ADRs in labeling, cautionary language on the limitations of comparing with other labels is usually included. For compounds in which there is a large impact from COVID-19, the cautionary language might need to be expanded to mention the potential for under- or over-reporting due to COVID-19. Furthermore, depending on the rarity of the

event and the extent of COVID-19 impact on the study, it is possible that it would be more appropriate to use a percentage from the non-COVID-impacted group or a percentage using patients who completed the trial prior to COVID-19 becoming a pandemic. When a lot of patients are unable to receive study medication for an extended period of time, creating an incidence rate that includes only events and time in which the patient was on study medication may be a good option. For events for which the impact of COVID-19 could persist for some time after COVID-19 logistical disruptions have passed, this EAIR could be limited to the time period before any COVID-19 disruptions. The best estimate to use for an event in a label may depend on the event and how it may have been impacted by COVID-19. As noted by Hemmings (2020), we surely do need to be attentive to the summary metrics used to quantify risks.

9. Summary of Safety Evaluations

Table 1 provides a summary of the recommendations that are applicable for most situations. However, details in previous sections are needed to fully understand the recommendations, possible exceptions, and cautionary notes. The analysis types included in this table are from select analyses described in the PHUSE white papers.

10. Concluding Remarks

The analytical approach to safety assessment for studies and submissions impacted by COVID-19 should remain focused on establishing the benefit/risk of the investigational product, staying as close to what was originally specified, and only making changes needed to address COVID-19 impact that would considerably alter the assessment of safety. For general assessment of AEs, SAEs, and AEs leading to permanent discontinuation of study drug (for controlled data), we believe the typical set of analyses will remain useful and should generally be carried out as planned, unless COVID-19 impact has been observed to be considerably different among treatment groups. If the impact due to COVID-19 is different across treatment groups, different or additional analyses should be considered. Furthermore, if the impact of COVID-19 is large, regardless of whether it is different across treatment groups, different methods of risk quantification may be needed for ADRs (see Section 8). For safety topics of interest expected to have a higher incidence due to a COVID-19 infection or due to quarantine or stay-at-home mandates (such as for depression), comparing exposure-adjusted incidence rates from uncontrolled data with a background rate from literature or alternative source could be more problematic than usual. For such safety topics, limiting data up to when COVID-19 became a pandemic or summarizing by COVID-19 subgroups (e.g., patients who had study visits impacted by COVID-19 vs.

Table 1. Summary of recommendations for common safety analyses for clinical trials impacted by COVID-19.

Analysis type	Recommendation for updating analysis plan
TEAEs, SAEs, AEs leading to study drug discontinuation—comparing percentages between treatments	General safety summaries, as normally created, are likely sufficient for detecting imbalances between treatments. If COVID-19 impact is different across treatment arms, different or additional analyses should be considered. If percentages are used for risk quantification, additional considerations are required.
TEAEs for uncontrolled data—Summarizing counts, percentages, exposure-adjusted incidence rates	General safety summaries, as normally created, are likely sufficient when used to identify rare events. When used for comparing with literature or a background rate, caution is required in interpretation, and additional analyses may be helpful.
Concomitant medications—Comparing percentages between treatments	The general summary table of concomitant medications, as normally created, remains useful for its intended purpose.
Labs/vitals—Comparing percentages between treatments (e.g., treatment-emergent highs and lows)	If data from local labs are included in the clinical trial database, low/normal/high should be determined using the local lab reference limits. If COVID-19 impact is different across treatment arms, different or additional analyses should be considered.
Labs/vitals—Simple summary statistics by visit	If COVID-19 impact includes a lot of missed visits, consider reporting means based on MMRM instead of the simple mean under the boxplot. If data from local labs are included in the clinical trial database, a decision is needed on which lab analytes can use combined data versus not or change to an alternative method that allows for the combination.
Labs/vitals—Comparing mean change to minimum/maximum values between treatments	If data from local labs are included in the clinical trial database, a decision is needed on which lab analytes can use combined data versus not or change to an alternative method that allows for the combination.
Hepatotoxicity—Plot of alanine aminotransferase versus total bilirubin	If local labs are used and data are not brought into the study database, there is a potential for missing patients that would otherwise have been in the Hy's law quadrant. Careful review of adverse event data would be required. If local labs are used and data are brought into the study database, use the limits from the local lab to determine the multiple of the upper limit of normal.
Intrinsic factors: Subgroup analyses for common TEAEs Case reviews	The general summary table or figure, as normally created, is likely sufficient. Include medication data as usual, and indication for use if collected. Include study drug exposure start/stop dates and information on visits impacted by COVID-19.
Safety topics of interest	For safety topics of interest that could have a higher incidence due to COVID-19 infection or due to quarantines, site closures, or economic hardship, summarizing by COVID-19 subgroups (e.g., patients without impact, patients with impact) or up to COVID-19 impact might be helpful for ADR decision-making.
Event characterization (e.g., event duration, transient versus persistent assessment)	If group summaries are planned, consider replacing with patient-based displays.
ADR communication (e.g., percentages to report) in product labeling	If the COVID-19 impact is large, for some ADRs, reporting the percentage from the non-COVID-impacted group might be warranted. Cautionary language may need to mention the potential for under- or over-reporting due to COVID-19. Using an exposure-adjusted incidence rate instead of a percentage may be useful.

patients who did not have any study visits impacted by COVID-19) should be considered. For analyses of laboratory measurements, analysis plans will likely need to be updated if there is a combination of measurements from local labs and central labs in the study database. When communicating ADRs in labeling, cautionary language on the limitations of comparing with other labels may need to be expanded to mention the potential for under- or over-reporting due to COVID-19 impact.

References

- Akacha, M., Branson, J., Bretz, F., Dharan, B., Gallo, P., Gathmann, I., Hemmings, R., Jones, J., Xi, D., and Zuber, E. (2020), "Challenges in Assessing the Impact of the COVID-19 Pandemic on the Integrity and Interpretability of Clinical Trials," *Statistics in Biopharmaceutical Research*, DOI: 10.1080/19466315.2020.1788984. [498,499]
- Allignol, A., Beyersmann, J., and Schmoor, C. (2016), "Statistical Issues in the Analysis of Adverse Events in Time-to-Event Data," *Pharmaceutical Statistics* 15, 297–305, DOI: 10.1002/pst.1739. [502]
- Allison, P. D. (2018), "For Causal Analysis of Competing Risks, Don't Use Fine & Gray's Subdistribution Method," available at <https://statisticalhorizons.com/for-causal-analysis-of-competing-risks>. [502]
- Association of Clinical Research Organizations (2020), "Considerations to Support Clinical Trial Monitoring Oversight During COVID-19," available at <https://www.acrohealth.org/wp-content/uploads/2020/03/ACRO-Statement-on-Monitoring-Oversight-FINAL-3.13.20.pdf>. [498]
- Bender, R., Beckmann, L., and Lange, S. (2016), "Biometrical Issues in the Analysis of Adverse Events Within the Benefit Assessment of Drugs," *Pharmaceutical Statistics*, 15, 292–296, DOI: 10.1002/pst.1740. [502]
- CDISC (2020), "CDISC Interim User Guide for COVID-19," available at <https://wiki.cdisc.org/display/COVID19/CDISC+Interim+User+Guide+for+COVID-19>. [499]
- CIOMS Working Groups III and V (1999), "Guidelines for Preparing Core Clinical Safety Information on Drugs," Council for International Organizations of Medical Sciences, Geneva. [499]
- CIOMS Working Group VI (2005), "Management of Safety Information From Clinical Trials," Council for International Organizations of Medical Sciences, Geneva. [499]
- Collins, S. H., and Levenson, M. S. (2020), "Comment on Statistical Issues and Recommendations for Clinical Trials Conducted During the COVID-19 Pandemic," *Statistics in Biopharmaceutical Research*, DOI: 10.1080/19466315.2020.1779123. [499]
- Crowe, B. J., Brueckner, A., Beasley, C., and Kulkarni, P. (2013), "Current Practices, Challenges, and Statistical Issues With Product Safety Labeling," *Statistics in Biopharmaceutical Research*, 5, 180–193, DOI: 10.1080/19466315.2013.791640. [500]
- Crowe, B. J., Xia, H. A., Berlin, J. A., Watson, D. J., Shi, H., Lin, S. L., Kuebler, J., Schriver, R. C., Santanello, N. C., Rochester, G., Porter, J. B., Oster, M., Mehrotra, D. V., Li, Z., King, E. C., Harpur, E. S., and Hall, D. B. (2009), "Recommendations for Safety Planning, Data Collection, Evaluation and Reporting During Drug, Biologic and Vaccine Development: A Report of the Safety Planning, Evaluation, and Reporting Team," *Clinical Trials*, 6, 430–440. [500]
- European Medicines Agency (2020a), "Guidance on the Management of Clinical Trials During the COVID-19 (Coronavirus) Pandemic," available at https://ec.europa.eu/health/sites/health/files/files/eudralex/vol-10/guidanceclinicaltrials_covid19_en.pdf. [498,499]
- (2020b), "Points to Consider on Implications of Coronavirus Disease (COVID-19) on Methodological Aspects of Ongoing Clinical Trials," Draft. [498,499]
- Foulkes, M. A. (2007), "Safety Assessment Versus Efficacy Assessment," in *Advances in Statistical Methods for the Health Sciences: Applications to Cancer and AIDS Studies, Genome Sequence Analysis, and Survival Analysis*, eds. J.-L. Auget, N. Balakrishnan, M. Mesbah, and G. Molenberghs, Birkhäuser Boston, pp. 323–334. [500]
- Geskus, R. B. (2016), *Data Analysis With Competing Risks and Intermediate States*, Boca Raton, FL: CRC Press. [502]
- Hemmings, R. (2020), "Under a Black Cloud Glimpsing a Silver Lining: Comment on Statistical Issues and Recommendations for Clinical Trials Conducted During the COVID-19 Pandemic," *Statistics in Biopharmaceutical Research*, DOI: 10.1080/19466315.2020.1785931. [499,503]
- Hengelbrock, J., Gillhaus, J., Kloss, S., and Leverkus, F. (2016), "Safety Data From Randomized Controlled Trials: Applying Models for Recurrent Events," *Pharmaceutical Statistics*, 15, 315–323, DOI: 10.1002/pst.1757. [502]
- Karvanen, J. (2003), "The Statistical Basis of Laboratory Data Normalization," *Drug Information Journal: DIJ/Drug Information Association*, 37, 101–107. [501]
- Kunz, C. U., Jörgens, S., Bretz, F., Stallard, N., Van Lancker, K., Xi, D., Zohar, S., Gerlinger, C., and Friede, T. (2020), "Clinical Trials Impacted by the COVID-19 Pandemic: Adaptive Designs to the Rescue?," *Statistics in Biopharmaceutical Research*, DOI: 10.1080/19466315.2020.1799857. [500]
- Ma, H., Ke, C., Jiang, Q., and Snapinn, S. (2015), "Statistical Considerations on the Evaluation of Imbalances of Adverse Events in Randomized Clinical Trials," *Therapeutic Innovation & Regulatory Science*, 49, 957–965, DOI: 10.1177/2168479015587363. [500]
- McDermott, M. M., and Newman, A. B. (2020), "Preserving Clinical Trial Integrity During the Coronavirus Pandemic," *JAMA*, DOI: 10.1001/jama.2020.4689. [498]
- MedDRA (2020), "MedDRA Version 23.0 Update 2020," available at <https://www.meddra.org/news-and-events/news/meddra-version-230-update-all-supported-languages-available>. [502]
- Meyer, R. D., Ratitch, B., Wolbers, M., Marchenko, O., Quan, H., Li, D., Fletcher, C., Li, X., Wright, D., and Shentu, Y. (2020), "Statistical Issues and Recommendations for Clinical Trials Conducted During the COVID-19 Pandemic," *Statistics in Biopharmaceutical Research*, DOI: 10.1080/19466315.2020.1779122. [498,499]
- Nautiyal, N., Rastogi, R., and Gamperl, H.-J. (2015), "Drug Safety Assessment in Clinical Trials: Concepts and Issues," *International Journal of Pharmaceutical Sciences and Research*, 6, 4159. [500]
- PHUSE (2013), "Analyses & Displays Associated With Measures of Central Tendency—Focus on Vital Sign, Electrocardiogram, & Laboratory Analyte Measurements in Phase 2-4 Clinical Trials & Integrated Submission Documents," Available at <https://www.phuse.eu/documents/working-groups/deliverables/analyses-displays-associated-with-measures-of-central-tendency-version-10-10-oct-13-11816.pdf>. [498,501]
- (2015), "Analyses and Displays Associated With Outliers or Shifts From Normal to Abnormal: Focus on Vital Signs, Electrocardiogram, and Laboratory Analyte Measurements in Phase 2-4 Clinical Trials and Integrated Summary Documents," available at <https://www.phuse.eu/documents/working-groups/deliverables/analyses-displays-associated-with-outliers-or-shifts-from-normal-to-abnormal-version-10-10-sept-15-11815.pdf>. [498]
- (2017), "Analysis and Displays Associated With Adverse Events: Focus on Adverse Events in Phase 2-4 Clinical Trials and Integrated Summary Documents," available at <https://www.phuse.eu/documents/working-groups/deliverables/adverse-events-white-paper-version-10-03-feb-17-11796-19835.pdf>. [498,500,501,502]
- (2019), "Safety Analytics Workshop Slides," available at <https://www.phuse.eu/documents/working-groups/deliverables/safetyanalyticsworkshop-PHUSEcss2019-21959.pptx>. [498,499,500,502]
- Proctor, T., and Schumacher, M. (2016), "Analysing Adverse Events by Time-to-Event Models: The CLEOPATRA Study," *Pharmaceutical Statistics*, 15, 306–314, DOI: 10.1002/pst.1758. [502]
- Rosario, L. A., Kropp, T. J., Wilson, S. E., and Cooper, C. K. (2012), "Join FDA/PhUSE Working Groups to Help Harness the Power of Computational Science," *Drug Information Journal*, 46, 523–524. [498]
- Senior, J. R. (2014), "Evolution of the Food and Drug Administration Approach to Liver Safety Assessment for New Drugs: Current Status and Challenges," *Drug Safety*, 37, 9–17, DOI: 10.1007/s40264-014-0182-7. [501]
- Singh, S., and Loke, Y. K. (2012), "Drug Safety Assessment in Clinical Trials: Methodological Challenges and Opportunities," *Trials*, 13, 138. [500]

- Stegherr, R., Beyersmann, J., Jehl, V., Rufibach, K., Leverkus, F., Schmoor, C., and Friede, T. (2019), "Survival Analysis for Adverse Events With Varying Follow-Up Times (SAVVY): Rationale and Statistical Concept of a Meta-Analytic Study," arXiv no. 1912.00263. [499,500]
- TransCelerate Biopharma Inc. (2020), "Clinical Study Report Considerations for Studies Disrupted by the COVID-19 Pandemic," available at http://transceleratebiopharmainc.com/wp-content/uploads/2020/05/TransCelerate_CSR_Considerations-for-Studies-Disrupted-by-the-COVID-19-Pandemic_May2020.pdf. [499]
- U.S. Food and Drug Administration (2009), "Guidance for Industry, Drug-Induced Liver Injury: Premarketing Clinical Evaluation," available at <https://www.fda.gov/media/116737/download>. [501]
- (2020), "FDA Guidance on Conduct of Clinical Trials of Medical Products During COVID-19 Public Health Emergency: Guidance for Industry, Investigators, and Institutional Review Boards," available at <https://www.fda.gov/media/136238/download>. [498,499]
- Unkel, S., Amiri, M., Benda, N., Beyersmann, J., Knoerzer, D., Kupas, K., Langer, F., Leverkus, F., Loos, A., Ose, C., Proctor, T., Schmoor, C., Schwenke, C., Skipka, G., Unnebrink, K., Voss, F., and Friede, T. (2019), "On Estimands and the Analysis of Adverse Events in the Presence of Varying Follow-Up Times Within the Benefit Assessment of Therapies," *Pharmaceutical Statistics*, 18, 166–183, DOI: 10.1002/pst.1915. [499,502]