1. Introduction

I am writing with not only the hope, but also the assumption, that the current COVID-19 pandemic caused by SARS-CoV-2 (coronavirus) infections will resolve, given time, and that both human health and our healthcare systems will eventually revert to states that are largely familiar, as they did after other major outbreaks of epidemic diseases. For sure there will be consequences, even once the pandemic is over, not least the likely reduction in healthcare budgets further complicating pricing negotiations for new medicines. There will be another virus circulating in humans, we might all receive another vaccination and there might be new treatments for COVID-19 available. On occasion these differences will present important considerations for identification of suitable conditions of use for a new medicine, but in respect of the development and use of most medicinal products any consequence should be negligible. If not, we should not only concern ourselves with the consequences for testing the candidate medicinal products that are currently in clinical trials, but also with whether the benefits and risks of authorized medicinal products as hitherto understood are significantly affected by such changes in their conditions of use.

In due course, clinical trials will be able to return to their former state (should we want them to). For now, however, the conduct of clinical trials is challenged in an unprecedented way, putting both participants and ongoing development programs at risk. At least for the latter, our discipline shares the responsibility to find solutions: to stop, to continue, or to amend studies that are currently ongoing and to make the best use of data for reliable estimation of treatment effects in trying to ensure that clinical trials can address their objectives. Unsurprisingly, the discipline has sprung to life resulting in contributions to publications from regulatory agencies and this Special Issue. I have had the pleasure to review the contribution from the Pharmaceutical Industry COVID-19 Biostatistics Working Group (Meyer et al. 2020), which offers a wide-ranging and well-structured set of considerations on actions to be taken at the level of an individual clinical trial that is affected by the COVID-19 pandemic.

2. Invoking ICH E9(R1)

Much of the structure to the discussion is given with reference to the ICH E9(R1) addendum on estimands and sensitivity analysis in clinical trials (ICH 2019), for which I was privileged to be the Rapporteur for the ICH Technical Working Group. Obviously, we did not have a global pandemic in mind when initiating or drafting the guidance. It is gratifying of course that the addendum is useful, but at the heart of that guidance was a simple premise that we had seemed to have forgotten in the excitement of discussing methods: simply put, that good decisions about design and analysis are best informed by clarity on the research question of interest. It is that simple premise which is now proving useful in discussing clinical trials impacted by the pandemic.

A related concept in the framework of E9(R1) was that the estimand can and should be defined prior to the design of a trial. Indeed, while it might seem academic, purist even, the definition of the treatment effect of interest exists outside, and without reference to, the experimental design that will be used to estimate it. For coronavirus-affected trials, Meyer et al. (2020) argue that the original primary objective should be maintained and, per the introduction above, I agree with that.

Even if an estimand was specified in the trial protocol it would certainly not have said “in a world without illness due to COVID-19.” However, that would have been the intent, that continues to reflect the clinical question of interest and the trial was designed for that purpose. However, intercurrent events (ICEs) will occur because of the COVID-19 pandemic. Perhaps drug supply is interrupted leading to treatment discontinuation, or the additional concomitant medications taken will affect the interpretation of measurements. Some trial participants may die. If, for example, a treatment-policy strategy reflected the clinical question of interest when the estimand was defined “in a world without illness due to COVID-19,” a treatment-policy strategy for ICEs that occur due to COVID-19 illness during the ongoing coronavirus pandemic would reflect a different clinical question. The consequence of retaining the original clinical question of interest is that the estimand on which the trial is based, and the analytic approach, will need to be refined.
The E9(R1) addendum states that changes to the estimand during the trial will be problematic, but the current pandemic must present an exception. Most clinical trials that are ongoing through the pandemic probably do not have an estimand pre-specified and the question arises whether any change to the analysis should be accompanied by the introduction of an estimand. Where methods are not aligned to an agreeable clinical question of interest this might be problematic for the sponsor. Nevertheless, it seems at least plausible that the E9(R1) framework will be used as a basis for discussion between sponsor and regulator and that introduction of an estimand would be encouraged as soon as practically possible.

3. Intercurrent Events

For trials that have specified an estimand, only a subset of ICEs might have been considered. For other potential ICEs there might be the unwritten assumption of “treatment-policy” or they might have been dismissed as being unlikely to occur, or too difficult to address. Omissions likely need to be addressed, and each ICE likely needs to be more granular, with the number at least doubled as being “due to coronavirus” and “not due to coronavirus.” The granularity of each ICE should also be considered in other regards: for how long treatment was interrupted due to the pandemic; the timing and extent of concomitant medications taken for coronavirus infection, etc. As discussed below, this is needed should different strategies be considered for framing the clinical question of interest. The additional granularity in the clinical question of interest might necessitate a change to the trial variable (or, conceivably, to the other estimand attributes). Specifically, an event that is a component of the existing primary variable might justifiably be considered an ICE. For example, if use of ventilator is a component of the primary variable, but the true clinical interest is on use of a ventilator not due to COVID-19, the variable and hence the estimand would be redefined, and a new ICE of ventilator use not due to COVID-19 would be introduced. Thereafter, appropriate choices for strategies can be discussed.

A first and perhaps greatest challenge is to collect information about the impact of the pandemic. Much of what is written here, and in Meyer et al. (2020), places additional weight on the data collected before the pandemic struck and some granularity in that definition also seems wise, if it is practical: perhaps specifying “pre-pandemic” cutoffs by study site rather than for the trial as a whole, or by alignment to the WHO declaration on March 11, 2020. For example, the date at which the first patient in a site became infected, the date at which study procedures at the site were first impacted, or the date at which local government “lockdowns” were first imposed. Beyond that, information is needed on the occurrence of ICEs and to support a judgment to be made on their classification. Operationally this must have already presented great challenges to trial sponsors: to plan and initiate additional data collection as the number of cases and consequences of infections and government lockdowns were increasing and peaking. Indeed, it seems likely that real-time collection might have been unfeasible in many trials and collection of sufficient information will need to be attempted retrospectively. In this context, sufficient information includes not only the occurrence and timing of ICEs, but clinical judgments: what was the cause of the patient dying, the ongoing serious adverse drug reaction or the coronavirus infection? Would the patient still have died without the infection? In the ICH E9(R1) Working Group we discussed at some length what was covered by “limitations to the data” under assessment of robustness in ICH E9 (1998). Perhaps the ability to correctly classify ICEs becomes one such limitation to which sensitivity analysis should be conducted.

4. Is It All Hypothetical?

If we think that the world post-pandemic, once illness due to coronavirus infection is again minimal, will indeed closely resemble the world pre-pandemic, and therefore that the original trial objectives remain appropriate, it seems logical to consider a “hypothetical” strategy for all ICEs that are related to the pandemic, whether due to operational constraints or individual patients becoming infected. Of course, this merits careful consideration both in terms of the appropriateness of the clinical question of interest and the possibility of reliable estimation. For the former I am comfortable, conceptually, with a hypothetical strategy regardless of the ICE and the clinical setting: whether considering discontinuations from treatment, taking of acetaminophen/paracetamol, oxygen (even hydroxychloroquine) as concomitant medication, or terminal events such as deaths. This includes settings with symptoms linked to those of COVID-19, for example, respiratory disease, or lung cancers. I have not thought of exceptions to this rule, perhaps others will. The latter question seems considerably more difficult and, per E9(R1), it should be agreed that reliable estimation is possible before the choice of estimand is finalized. As above, can the ICE be correctly attributed as due to coronavirus or not? What is the extent of information available on affected patients (predicting week 24 values for a patient with measurements up to week 20 seems to be a meaningful exercise whereas predicting outcomes for a patient with measurements only at baseline does not)? Is the disease sufficiently well understood that the development of a model to make predictions is reasonable? Most importantly perhaps, is the remainder of the dataset sufficiently well populated to serve as a basis for making predictions? It seems likely that some trials will need to be considered futile because the pandemic has struck before the database has become sufficiently mature.

E9(R1) discussed different hypothetical strategies, distinguishing between the effect of a treatment under different conditions from those of the trial that can be carried out and a hypothetical scenario in which ICEs would not occur, based on differences in the clinical and regulatory interest in the clinical questions associated with those. This distinction is important in relation to ICEs that occur because of treatment but, if the thinking outlined above is followed, is unimportant (and perhaps only semantic) for ICEs due to the pandemic. However, in agreement with Meyer et al. (2020), the conceptual intent would not be to predict “what would have happened in patients whose treatment or condition was affected by the pandemic had remained alive, on treatment and without concomitant treatments affecting measures of outcome.” Some affected patients
would still have experienced lack of efficacy or toxicity and discontinued treatment, some might still have died.

If the hypothetical strategy is agreed for all ICEs that are related to the pandemic, and an estimand is constructed accordingly, analyses to investigate whether estimates of treatment effect would have been similar under other strategies should not be requested, but the sensitivity analysis to check the performance of the statistical model in making predictions will be extensive. Returning to the granularity with which each ICE is defined, a hypothetical strategy for all ICEs due to coronavirus will place a greater burden on the dataset for making predictions of outcomes. The extent of sensitivity analysis will be increased as will the risk that results are found sensitive to the assumptions in the analytic approach. A more judicious specification of ICEs, selecting “treatment-policy” for relatively unimportant breaks in treatment administration due to pandemic or the use of certain concomitant medications, even if there is some potential for measure of outcome to be affected, might provide a better balance and put less stress on the available data for making predictions.

5. When Is Data Being Missing Not Missing Data?

A key concept in the E9(R1) guideline is to distinguish between missing data and data points after ICEs that are agreed not to be relevant to the research question. ICEs should be reflected in the estimand, whereas missing data are defined as data that would be meaningful for the analysis of a given estimand but were not collected. Because of the pandemic, extensive amounts of data that were initially planned for collection will not have been collected, some might not even exist. In an approach aligned with the hypothetical, and in respect of ICEs due to coronavirus, the mass closure of sites and the difficulties or reluctance of patients traveling for dosing and assessments does not necessarily present a huge missing data problem but instead presents a problem for prediction under a hypothetical. When it comes to the choice of analytic approach, the distinction might seem unimportant since the same analytic approach might be selected to address both issues in the main analysis. A clear understanding is important, however, for continued data collection and prioritizing efforts to collect information that is needed for estimation, and for sensitivity analysis.

As discussed by Meyer et al. (2020), the pattern of pandemic-related ICEs and missingness should be investigated and the main or sensitivity analysis may include additional baseline covariates (e.g., age and co-morbidities) and even post-randomization outcomes. I was struck by this attention to patterns in the types of patients experiencing ICEs or with missing data. Quantification and impact assessment of the pattern of missingness and the adequacy of the statistical model in that regard is notable by its absence from our usual clinical trial analysis and reporting, most of which relies (wrongly in my opinion) on the assumption that MAR is valid based on the few selected covariates that are included in the statistical model.

Another discussion within the E9(R1) working group was the extent to which the addendum should reflect potential regulatory preferences for choice of strategy. In large part this was avoided, but a piece of pragmatic guidance was that:

Characterizing beneficial effects using estimands based on the treatment-policy strategy might also be more generally acceptable to support regulatory decision making, specifically in settings where estimands based on alternative strategies might be considered of greater clinical interest, but main and sensitivity estimators cannot be identified that are agreed to support a reliable estimate or robust inference.

For trial objectives related to a demonstration of superiority for the experimental agent, use of data and methods aligned to the treatment-policy strategy might still provide estimates of a treatment effect that, before E9(R1), would (imprecisely) have been regarded as conservative and might serve as a suitable basis for decision making. As ever, non-inferiority studies remain more complicated. For those research questions, the regulator has the additional consideration of whether the trial is sensitive to detect any difference between treatments. A treatment-policy approach to, for example, an interruption of drug supply affecting both treatment groups should not be found acceptable.

6. Other Impacts

Other issues will arise. Sponsors are obliged to consider not only patient safety in ongoing trials, but also whether continuation is futile and whether amendments are needed. However, a look into the data of an ongoing trial with the possibility of making amendments would usually draw criticism. Many trials are conducted open-label or, by choice or by necessity, as single-arm trials. Amendments made, or suspected as having been made, based on patterns in the emerging non-affected trial data rather than based on the impact of the pandemic might rightly attract criticism. It will not always be easy to establish what was the motivation for a change. Hence the calls for an active role for data monitoring committees (DMCs). To the extent possible, sponsors should try to keep their investigations related to the impact of the pandemic, prespecified, documented, and auditable (who knew what, and when) and, where possible, made by bodies independent from day-to-day study conduct.

When there has been an interim look at a database, or a modification to trial conduct or the plan for analysis, a check for consistency of effects pre- and post-modification is common. The draft points to consider on implications of coronavirus disease (COVID-19) on methodological aspects of ongoing clinical trials from the European Medicines Agency (https://www.ema.europa.eu/en/documents/scientific-guideline/points-consider-implications-coronavirus-disease-covid-19-methodological-aspects-ongoing-clinical_en.pdf) indicates a role for DMC investigation of “…the impact of the three phases (pre, during, and post coronavirus) to understand the treatment effect as estimated in the trial.” Given that the treatment and assessment of an individual patient might span two or even all three pandemic phases, the methodological approaches to these investigations will not be straightforward. Guidance on the objectives for these investigations and for related assessments of consistency would be welcome. It is understandable that assessment of the treatment effect pre-pandemic might give useful insights to a reviewer. However, where sites and patients are affected by the pandemic, the experiment itself is not “consistent” over these phases, so we might not expect the
effects of treatment to be. Again, in which effect of treatment (estimand) are we interested? Consistency of treatment effects does not seem meaningful beyond of effects based on the hypothetical strategy for ICEs due to coronavirus, and that seems a trivial target to achieve if the data before the pandemic are used to predict those that are considered irrelevant to the research question because of the pandemic.

Unfortunately, some trials will fail. Consider a trial that fails its primary hypothesis test. This might be despite best efforts for continued data collection or might be where continued data collection is deemed impractical or irrelevant and the trial is terminated early. The sponsor's analysis appears to confirm that the effect size was large, perhaps larger even than was anticipated. The sponsor has an argument to make that the trial only failed due to the pandemic. Perhaps the product addresses an unmet need, providing a treatment option where one currently does not exist, or an important improvement on available therapies. Supportive approaches might look to integrate data from external sources, to augment the control arm, or pool trial data with results from previous studies but regulators will have to consider approvals based on a greater than usual degree of uncertainty and use relevant and feasible post-authorization data generation to complement the pre-authorization study or studies. In the European Union, for example, a broader use of Conditional Marketing Authorization might be foreseen. The increased risk of incorrectly authorizing an ineffective treatment might have to be accepted in that scenario, since the risks to public health of denying access to truly effective treatments would seem greater.

Although both efficacy and safety have come to be scrutinized, assessment of safety and informing prescribers and patients on risks for toxicity are key responsibilities for regulators. Despite that, I have been guilty here—aren't we all—of a focus in thinking and language on efficacy. When discussing E9(R1), we met with the proposal that efficacy should be appraised under ideal conditions, whereas safety should be appraised as observed. The argument went that we risk double counting toxicities: that serious toxicity or tolerability issues leading to discontinuation are already counted in the safety database, and hence discontinuations due to adverse events should not also count against efficacy. I think I understood the argument, but this never made good sense to me. My general preference for treatment-policy in respect of adherence to treatment (at least for superiority settings) is based on trying to get externally valid estimates of both beneficial and harmful drug effects. Specifically, unless it can be explained why adherence would be different in practice than it would in the trial, as I think it can in relation to the pandemic, patients being unable to continue with treatment for whatever reason should be reflected in estimand for efficacy. A somewhat tongue in cheek response was that, if you want to use the “bad” hypothetical for efficacy, you should also make predictions for what additional toxicities would have been observed had patients been able to continue with treatment. Now the door to a “good” hypothetical for efficacy is open, what additional toxicities would have been reported in the safety database had some patients been able to continue with treatment? The missing data (forgive me, you know what I mean) in respect of safety still has to be considered and making predictions for toxicities from a statistical model seems difficult and conceptually undesirable. In all likelihood, companies and regulators will have to work with the safety data that have been collected whilst treatment has been taken. Already trials are sized based on a primary efficacy variable and assessment of safety seems likely to be compromised as safety databases are reduced in respect of numbers of patients and exposure times. Simple incidence rates and exposure-adjusted rates are both used and misused in current practice and regulators will surely have to be attentive to the summary metrics used to quantify the risks of adverse events from short-term and from long-term treatment. Randomized comparisons can certainly be important for assessment of safety, but perhaps deficiencies in a safety database are more easily remedied than those for efficacy. Data from use of the product in other trials, even other indications can be relevant and additional cohorts of patients can be initiated to provide greater precision and to obtain adequate long-term exposure. This might be a useful topic for consultations with regulators on mitigating the impact of the pandemic on the evidence base for future regulatory decisions.

7. A Silver Lining?

It seems difficult to construct general guidance on the scenarios where it is useful to attempt ongoing data collection. Short dosing regimens for vaccines and one-off dosing for gene therapies seem to present different problems to treatments dosed chronically, but for each of those different variables and frequencies of assessment will be important. Some studies might essentially become pragmatic trials, recording treatment assignment and outcome (e.g., mortal status) and that might still be worthwhile. Other studies, where key trial data are reliant, for example, on blood draws, functional tests to be conducted under specific conditions, administration of product by a healthcare professional present greater challenges, though even for these there might be options for local assessment, for example, through primary care rather than a tertiary care or academic trial site. However, if it is agreed that continuation of the trial remains appropriate, how might this be done? Well, necessity is the mother of invention and herein lies our silver lining. Over recent months across many occupations, more people have been working remotely, not only from choice but out of necessity. Similarly, sites conducting clinical trials have been closed, or overwhelmed with other duties, and trialists are forced to consider remote dosing and remote (or self-) assessment by patients for clinical trial variables, and regulators will be forced to consider the validity of these other methods for dosing, and for generation and collection of data. There is nothing like a real-life testbed for novel approaches to overcome inertia. Lessons learned will be useful in progressing conversations on “decentralized” or “hybrid” trials where the patient's home effectively becomes in part or in whole a trial site. Remote monitoring for data quality and an increase for centralized statistical monitoring might reduce or replace site visits, with a focus on the quality and completeness of data that will be key to the regulator and other stakeholders understanding drug effects. Might regulators even attempt remote GCP inspections?

The potential for learning about operational efficiencies is interesting in the context of ongoing discussions on late-phase
clinical trials and the search for efficiencies in terms of methodology, including use indirect/external comparisons, augmented control arms, informative integration of earlier phase results, and use of data that have been generated and collected in clinical practice. Each of these comes with a risk to the internal validity of a trial to an extent that more efficient operational trial conduct might not. The true benefits of late phase clinical trials should be the possibility to understand the effects of an experimental treatment by randomization of patients to treatment conditions, preferably in a blinded manner, under conditions that are well understood (but not necessarily restrictive or artificial), with prespecified approaches to the variables that will be collected, the handling of those data and their analysis. Motivations for use of novel methodological approaches have been the high cost and time-consuming nature of clinical trial setup and conduct, or their lack of external validity. Such motivations seem flawed, or at least incomplete. Costs and time are driven primarily by operations, not by the fundamental principles of experimental design. To address challenges, commentators, industry and occasionally regulators have shown willingness to compromise those fundamental principles of clinical trials. Patients deserve good drugs. They also deserve an assurance that drugs have positive benefit-risk and an accurate description of their potential treatment effects. So, whilst efforts to expedite drug development should be made, the randomized controlled trial, in some form, should remain the basis for understanding treatment effects. It seems more attractive to try to reduce site and monitoring costs and to improve external validity through advances in clinical trial operations than through greater use of methods that present a risk of compromising experimental design.

8. Next Steps
We all look forward to receiving additional regulatory guidance for the ongoing conduct and eventual analysis of clinical trials affected by the pandemic. Guidance to cover all questions related to all scenarios cannot be expected, but some principles would be helpful to guide product-specific discussions and to dictate best practice for how coronavirus-related investigations and amendments should be documented. Regulatory guidance on how to document such information systematically would seem helpful and might well be in preparation. However, a focused and rapid work through International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) or another global forum seems preferable to multiple regional guidances being issued. A summary of clinical trial actions related to coronavirus might be a useful supplement to the ICH Common Technical Document (ICH M4 2003) perhaps as an annex to Module 2.5 or a stand-alone document in Module 2.7, to be prepared consistently across companies and development programs. This might detail the trials in the application that were affected, when different trial sites were affected, the states of the databases “pre-pandemic,” what investigations were performed mid-trial to consider impact, amendments in plans for analysis, the audit trail for access to trial data etc.

9. Conclusion
The coronavirus pandemic is an unprecedented challenge for ongoing clinical trials, and the paper from the Pharmaceutical Industry COVID-19 Biostatistics Working Group is a well-considered reflection that can be welcomed by trial sponsors and regulators alike.

References
ICH (2019), “Addendum on Estimands and Sensitivity Analysis in Clinical Trials to The Guideline on Statistical Principles for Clinical Trials,” available at https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf. [414]
ICH E9 (1998), “Statistical Principles for Clinical Trials,” available at https://database.ich.org/sites/default/files/E9_Guideline.pdf. [415]
ICH M4 (2003) “Common Technical Document,” available at https://www.ich.org/page/ctd. [418]
Meyer, R. D., Ratitch, B., Wolbers, M., Marchenko, O., Quan, H., Li, D., Fletcher, C., Li, X., Wright, D., Shentu, Y., Englert, S., Shen, W., Dey, J., Liu, T., Zhou, M., Bohidar, N., Zhao, P-L., and Hale, M. (2020), “Statistical Issues and Recommendations for Clinical Trials Conducted During the COVID-19 Pandemic,” Statistics in Biopharmaceutical Research. DOI: 10.1080/19466315.2020.1779122. [414,415,416]