Comments on
ICH E20 Draft Guidance
Adaptive Designs for Clinical Trials

presented by

Vance W. Berger, PhD — National Institute of Health, USA
Kerstine Carter — Boehringer Ingelheim Pharmaceuticals, Inc., Ridgefield, CT, USA
Peter Jacko, PhD — Lancaster University, UK
Johannes Krisam, PhD — Boehringer Ingelheim Pharma GmbH & Co. KG, Ingelheim, Germany
Ayon Mukherjee, PhD — Population Health Sciences Institute, Newcastle University, UK
Lukas Pin — MRC Biostatistics Unit, University of Cambridge, UK
David S. Robertson, PhD — MRC Biostatistics Unit, University of Cambridge, UK
Oleksandr Sverdlov, PhD — Novartis Pharmaceuticals Corporation, East Hanover, NJ, USA
Diane Uschner, PhD — F. Hoffmann-La Roche, Basel, Switzerland
Sofía S. Villar, PhD — MRC Biostatistics Unit, University of Cambridge, UK
Wenle Zhao, PhD — Medical University of South Carolina, USA

listed in alphabetical order, on behalf of the

Randomization Working Group of the ASA Biopharmaceutical Section, PSI/EFSPI Special Interest Group, sub-team of the DahShu Innovative Design Scientific Working Group (IDSWG).

1 December 2025

1 General Comments

We welcome this guidance centered on good practices for design and analysis for adaptive trials. Strikingly, randomization, a cornerstone of randomized clinical trials, receives only peripheral mention within ICH E20. Randomization and its rigorous implementation are not merely a statistical detail; they are essential for trial integrity, especially in adaptive settings. Proper randomization is a cornerstone of confirmatory clinical trials, ensuring internal validity and unbiased results. The dynamic nature of adaptive designs inherently strains traditional randomization methods. This is evident in Response-Adaptive Randomization (RAR), but equally important in Multi-Arm Multi-Stage (MAMS) or platform trials that involve dropping or adding arms, which necessitate dynamically changing allocation ratios.

Randomization is a key design element and must be better described as such throughout the guidance. Crucially, the current guidance fails to convey the fundamental challenge that adaptation itself poses to the proper implementation of randomization beyond the simple acknowledgment of response-adaptive allocation. This operational complexity requires clear planning and rigorous procedures to maintain trial integrity, a crucial area that must be addressed to ensure the robustness and validity of the resulting trial designs.

We appreciate the guideline's dedication to trial integrity and its recognition of associated risks and mitigation measures within adaptive designs. However, we feel some particularly critical risks associated with the choice of randomization procedure merit a more explicit mention. Specifically, our first comment deals with the potential for trial personnel to predict or guess subsequent treatment assignments in trials with no blinding, which is a direct cause of selection and ascertainment

bias. While this falls within trial integrity and operational risks, an explicit mention in reference to randomization procedure choice would strengthen the guidance.

2 Re: Section 3.5 "Maintenance of Trial Integrity"

General Comments

We appreciate that the guidance dedicates an entire section to trial integrity, and acknowledge that risks for trial integrity and associated mitigation measures within adaptive trial designs are mentioned. However, we feel that one particularly important risk was missed in this section: the potential for trial personnel trying to predict or guess the subsequent treatment assignments, which can cause several kinds of bias. Especially for open-label trials, which are explicitly mentioned in lines 306-311 to be particularly sensitive to breaches of trial integrity, we feel that the predictability of treatment assignments poses a severe risk to trial integrity which could be addressed in this guideline as well. The permuted block design, which is still the most frequently used randomization method throughout all clinical trials, is known to have a very high proportion of deterministic assignments, thus being very vulnerable to attempts of investigators trying to guess the subsequent treatment assignment (Berger et al. 2021). Notably, the guidance addresses the problem of selection bias arising from predictable deterministic assignments in lines 556-560, but only in the context of non-randomized fully deterministic response-adaptive allocation procedures. It does not acknowledge that the most widely used randomization method - permuted block randomization - shares the same vulnerability. We believe that a section dedicated to trial integrity within this guideline offers an opportunity to address the shortcomings of the permuted block design especially with respect to open-label trials, and could list other randomization methods as alternatives. The class of maximum tolerated imbalance (MTI) procedures achieve the same degree of control of the maximal imbalance of a randomization sequence, while providing a procedure that is more random, thereby less predictable, hence reducing the risk of selection bias. Examples for these procedures are, e.g. the Big Stick Design (Soares & Wu, 1983), the Biased Coin Design with Imbalance Tolerance (Chen 1999), the Block Urn Design (Zhao & Weng 2011), or the Maximal Procedure (Berger et al. 2003).

In addition, we feel that the guideline also should contain an explicit statement to limit access to an open-label trial database, thereby reducing the risk of biases introduced by knowledge of the sponsor. As an example, Higgins et al. (2025) recommend that the "sponsor statistician should be blinded, that is, not have the knowledge of subjects' assignments, until the database is locked and the study is officially unblinded.". In this context, using MTI procedures will also represent an important measure, as this will decrease the risk that the sponsor might be able to predict subsequent treatments based on the current history of treatment assignments within the database. In addition, some recommendation on how to best handle open-label databases, including potential risk mitigation measures, such as restricting parts of the database containing information on the treatment of the patients (or information to deduce the treatment), the use of mock or scrambled information, etc. would be worthwhile additions to this section.

General Recommendations

We suggest that the section on "Maintenance of Trial Integrity" should explicitly address the risk of trial personnel predicting future treatment assignments, which can introduce bias, particularly

in open-label trials. The guideline could acknowledge the limitations of the widely used permuted block design, which is highly predictable, and recommend alternative randomization methods from the class of Maximum Tolerated Imbalance (MTI) procedures. Additionally, we suggest that the guideline might benefit from including recommendations to restrict access to open-label trial databases to prevent sponsor-related bias, for example by blinding sponsor statisticians until database lock and implementing measures such as limiting access to treatment-related fields or using scrambled data. Combining these measures with less predictable randomization methods would significantly strengthen trial integrity.

References

- Berger VW, Bour L, Carter K et al. (2021). A roadmap to using randomization in clinical trials. BMC Medical Research Methodology 21(1):168.
  doi: 0.1186/s12874-021-01303-z.

- Berger VW, Ivanova A, Knoll MD (2003). Minimizing predictability while retaining balance through the use of less restrictive randomization procedures. Statistics in Medicine 22(19):3017-3028.
  doi: 10.1002/sim.1538.

- Chen YP (1999). Biased coin design with imbalance tolerance. Communications in Statistics. Stochastic Models 15(5):953-975.
  doi: 10.1080/15326349908807570.

- Higgins KM, Levin G, Busch R (2024). Considerations for open-label randomized clinical trials: Design, conduct, and analysis. Clinical Trials 21(6):681-688.
  doi: 10.1177/17407745241244788.

- Soares JF, Wu CFJ (1983). Some Restricted randomization rules in sequential designs. Communications in Statistics - Theory and Methods 12(17):2017-2034.
  doi: 10.1080/03610928308828586.

- Zhao W, Weng Y (2011). Block urn design - a new randomization algorithm for sequential trials with two or more treatments and balanced or unbalanced allocation. Contemporary Clinical Trials 32(6):953-61.
  doi: 10.1016/j.cct.2011.08.004.

3 Re: Section 4.5 "Adaptation to Participant Allocation"

General Comments

The section addressing changes to treatment allocation is the portion of the guidance that most explicitly relates to randomization. However, this text is problematic because it is not only unduly conservative regarding the benefit-risk profile of Response-Adaptive Randomization (RAR), but it also displays an apparent lack of awareness of current efforts in the statistical literature to fully address the stated concerns. This deficit leads to the re-encapsulation of several long-standing myths and misconceptions; for instance, the guidance fails to mention frequentist optimal RAR

methods, which would be possibly more suitable for the confirmatory setting than other approaches that have been used in past trials, and it appears to make an unnecessary, covert mention of the Play-the-Winner design and the infamous ECMO trial. This failure to engage with contemporary solutions risks unnecessarily discouraging the use of scientifically sound and statistically and/or ethically superior randomization strategies.

In general, the current text contributes to the general confusion between (target) "allocation ratios" between per-group sample sizes and per-participant "randomization ratios/probabilities", which are distinct concepts, as can be very clearly seen in platform trials, where one can modify randomization probabilities to speed up enrollment to better-performing treatments while still eventually reaching the originally planned allocation ratios and per-group sample sizes.

We understand that the guidance is typically designed to target the organizations who may not plan trials appropriately. However, overly restrictive language can influence non-statistical stakeholders (e.g., clinicians) and lead to rejection of innovative designs. For the particular case of RAR, it would be helpful if regulators identified circumstances where its use is most compelling (e.g. rare indications with expectation of transformative efficacy) where alternative trial designs may not be feasible.

General Recommendations

1. The guidance should specify the situations in which response adaptive randomization (RAR) approaches are warranted and discuss appropriate statistical methods for RAR trials. For example, RAR may be particularly appropriate for indications for very rare diseases where there are expectations of strong efficacy benefits. In these situations, the aim would be to establish statistically significant superiority while maximizing the number of treatment successes in the trial. Unbiased estimation, while important, may not be the primary goal.

2. Bayesian inference can be highly beneficial for RAR trials, especially in settings with limited recruitment (e.g., pediatric populations, where adult data can inform pediatric estimates). The use of Thompson sampling is particularly beneficial in such cases. In this section, we recommend discussing Bayesian inference and Thompson Sampling explicitly, as opposed to focusing solely on combination tests, which may lose efficiency due to suboptimal planning of stage-wise weighting.

3. The guidance should distinguish between RAR applied only among investigational treatments while protecting the allocation to the control arm versus RAR including the control arm as well. Such distinction is crucial as it remarkably affects the performance and operating characteristics of the trial, as evidenced in a number of papers. Also, we recommend that the difference between trials using a fixed total sample size compared to trials with fixed per-group sample sizes should be explained.

4. We recommend this section to include randomization approaches other than RAR that might be appropriate for adaptive design trials. For example, the section mentions covariate-adaptive randomization (CAR) approaches where probability of treatment allocation depends on accumulating covariate information, covariate of the incoming patient, and the treatment allocation history. We suggest that the guidance should provide examples to illustrate when CAR would be more appropriate than stratified randomization, and to illustrate what measure should be taken to control the Type I error rate when implementing CAR.

5. The guidance should also include recommendations on the suitability of the use of frequentist optimal allocation proportions while implementing RAR in practice, and incorporate strategies for achieving optimal allocations (maximizing power or minimizing the total sample size), such as Neyman allocation or the RSIHR allocation approach (see references below).

6. When covariate information is available and RAR is beneficial for patients (specifically for rare disease trial settings), Covariate-Adjusted Response-Adaptive (CARA) designs should be considered. The guidance should include discussion of these designs and appropriate use cases.

Specific Comments and Recommendations

- **Line 514. Comment:** The heading "Adaptation to Participant Allocation", suggests to discuss randomization methods that depend on the participant allocation to each arm and aim at somehow balancing that to attain some objective. However, this topic is not currently addressed in the section.

  Recommendation: We recommend to split Section 4.5 into Section 4.5a "Adaptations Based on Observed Participant Allocation" and Section 4.5b "Adaptations Based on Observed Participant Responses" and discuss these two topics in clearly separated parts. The former would discuss "allocation-adaptive" and "covariate-adaptive" approaches, which aim at balancing the per-group sample sizes and per-subpopulation sample sizes, while the latter would discuss "response-adaptive" (which could be blinded or unblinded) approaches. In fact, these two section titles are more general, and current sections 4.1, 4.2, 4.3, and 4.4 could be moved into one of these two categories. For instance, early trial stopping, or arm dropping in multi-arm trials, are clearly response-adaptive methods.

- **Line 515. Comment:** The sentence "In a randomized trial, participants are typically allocated to treatment arms according to fixed randomization probabilities." is not true.

  In the context of randomization methodology, it is essential to distinguish among the following concepts

  - **targeted allocation probabilities** derived from targeted allocation ratios (which would be fixed probabilities of 0.5 for each arm in a randomized trial with a 1:1 target allocation ratio)
  - **conditional allocation probabilities**, i.e. the probabilities for a given patient to receive a given treatment conditional on the previous treatment assignments – these are not constant by design for any restricted randomization procedure, but are changed in order to meet some balance prerequisite (i.e. are set to 0 or 1 at the end of each block within a permuted block design)
  - **unconditional allocation probabilities**, being the probabilities for a given patient to receive a given treatment unconditional on the previous assignments – these probabilities are generally constant for trials with equal allocation and therefore typically coincide with the targeted allocation probabilities, but are also known to vary under several procedures with unequal allocation ratio, such as a naïve extension of the biased coin design to unequal allocation, or unequal allocation minimization (Kuznetsova Tymofyeyev 2012), thereby potentially causing several types of bias.

For illustration, consider a conventional (non-adaptive) confirmatory trial using permuted block randomization with a fixed 1:1 target allocation ratio. While the targeted allocation probability remains 0.5 for each arm, the conditional probabilities within each block vary between the assignments within the block. For instance, with a block size of 4 and a permutation of AABB:

- For the first participant, the conditional and targeted probabilities coincide (0.5 for both Arm A and B).
- For the second participant, given the first was A, the conditional allocation probabilities become 1/3 for A and 2/3 for B.
- For the third and fourth participants, allocation becomes deterministic (resulting in conditional allocation probabilities of 0 for A, 1 for B).

In fact, only under simple randomization (often-times also called complete or unrestricted randomization), where each assignment is independent and corresponds to a fair coin toss, the targeted, conditional, and unconditional allocation probabilities coincide at every step. This also holds true for generalizations of simple randomization to unequal target allocation ratios and/or more than two treatment arms.

Nevertheless, regardless of the targeted allocation ratio, the number of treatment arms, or whether RAR or a design with fixed target allocation probabilities is used, controlling imbalance is typically considered important within most clinical trials. For example, the ICH E9 guidance (ICH, 1998) states that "Although unrestricted randomisation is an acceptable approach, some advantages can generally be gained by randomising subjects in blocks", and further explains that imbalance restrictions help protect against bias coming from time trends and ensure approximately equal group sizes. This feature is not unique to permuted block randomization, it can also be achieved by the broader class of MTI randomization procedures (see above comment on Section 3.5 of ICH E20). In any case, achieving balance control in the sense of ensuring that the observed allocation ratio between treatment arms is sufficiently close to the targeted allocation ratio can only be guaranteed by randomization methods that adapt the conditional allocation probabilities so that they become different from the targeted allocation probabilities (as outlined in the example above).

Recommendation: Remove that sentence, or replace by the explanation as above, or rephrase as "In conventional (non-adaptive) confirmatory trials, participants are typically allocated to treatment arms according to a fixed target allocation ratio.". This applies to both Section 4.5a and Section 4.5b, but it may not be the best opening sentence/paragraph (see the next comment).

- **Line 515. Comment:** The opening sentence of Section 4.5b on "Adaptations to Observed Participant Responses" should introduce a setting, a problem or a trial objective that might need to be addressed, like the opening sentences in Sections 4.1, 4.2, 4.3, and 4.4 do.
[4.1. "During the conduct of a clinical trial, accruing data can provide information that makes it no longer appropriate to continue the trial." 4.2. "Even after a carefully planned and conducted early-phase development program, a considerable degree of uncertainty might exist in the parameter assumptions that affect the sample size calculations for a clinical trial." 4.3 "In certain settings, there may be remaining uncertainty about the patient population

who should be treated with a new treatment." 4.4. "Some trials are conducted with the intent to evaluate more than one treatment. The multiple treatments might be different drugs or different doses of a single drug."]

Recommendation: Replace the opening sentence by the following paragraph: "Even after a carefully planned and conducted early-phase development program, a considerable degree of uncertainty might exist in the parameter assumptions that affect the choice of the target allocation ratio between a treatment and the corresponding control arms, such as the nuisance parameters mentioned in Section 4.2. Similarly, trials conducted with the intent to evaluate more than one treatment such as those discussed in Section 4.4 may aim at identifying the better treatment(s) faster in order to bring them to the out-of-trial population as early as possible. During the conduct of a clinical trial, accruing data on participants' outcomes can provide information, as discussed in Section 4.1., that makes it reasonable and desirable to adjust the target allocation ratio(s). Like in Section 4.3, there may be remaining uncertainty about the patient population who should be treated with a new treatment, and accruing data on participants' outcomes together with observed covariates may make it reasonable to adjust the target allocation ratio(s) for different patient subpopulations." Then, instead of "Alternatively, there are..." continue with "There are..." and the rest of the current opening paragraph, except for the last sentence (see the next comment).

- **Lines 523-525. Comment:** The guidance states that the key idea for RAR approaches "is to assign new participants with greater probability to treatment arms that have had, to that point, more positive outcomes than to other treatment arms." Improving expected outcomes of the trial participants is only one of the possible goals of RAR, and probably not the leading one. Other objectives include improving power to detect a difference between the treatment groups, reducing the total number of patients in the trial, or speeding up enrollment to better treatments.

  Recommendation: In Section 4.5b, add examples of key goals of response adaptive randomization, as mentioned above.

- **Line 526. Comment:** There is a lot of confusion and myths in the academic literature and among trial statisticians about what RAR is. It is necessary to somehow explain and define it. The guidance barely touches on that.

  Recommendation: In Section 4.5b, add the following new paragraph: "There is a broad palette of methods for randomization of trial participants that fall under the general umbrella of response-adaptive randomization (RAR). Adapting the target allocation ratio for the next block of participants is a softer version of the more aggressive adaptations discussed in the previous subsections such as early stopping of a trial (or of a study within a master protocol trial), treatment dropping, modification of the total sample size of a trial (or of a study within a master protocol trial), or subpopulation dropping. RAR can be specified to temporarily decrease the randomization probability to a particular treatment in order to lower the number of participants from the next block on it, and such decrease might even mean zero-ing it, i.e., temporarily dropping it. The block size can be defined by update milestones, which can be specified in terms of the number of participants allocated or observed (like in the permuted block randomization), or specified in calendar time. At the next update, that treatment has a chance to recover if the new set of outcomes data suggests that the previous adaptation should

be corrected because it was caused by random variability. Such a correction is not possible with the more aggressive adaptations such as the (permanent) dropping of a treatment. Such updates of the target allocation ratio are typically specified algorithmically, and do not need to be approved by the IDMC for every occasion, although they should be pre-specified and described in a specific document rather than the protocol, for instance, in a confidential appendix to the IDMC charter, as discussed in Section 3.5, to facilitate the evaluation of trial operating characteristics (e.g., expected sample size and power) and ensure that the IDMC understands and is in agreement with the anticipated adaptation rule. How exactly the RAR adapts the target allocation ratio depends on the objective(s) set for the particular trial, mentioned in the previous paragraph." Then continue, as a new paragraph, with the one starting with "RAR is sometimes valued.."

- **Lines 529-534. Comment:** "RAR designs are susceptible to bias and inflation of the Type I error probability in the presence of overall time trends. For example, a RAR design would more likely show a false positive treatment effect if earlier-enrolled participants are both more likely to be assigned to control and to have a poor prognosis (e.g., because of changes in background care or participant characteristics over time) than later-enrolled participants."

  Recommendation: We recommend removing this text because it suggests that RAR could increasingly allocate fewer subjects to the control arm over time, assuming treatment arms are more effective. Instead, there are well established approaches where RAR studies adjust randomization ratios among active arms but hold the control arm allocation constant throughout the trial to enable sufficient sample size for treatment versus the control arm to have sufficient power for treatment comparison. This applies both to trials comparing multiple treatments, where RAR can be implemented only among the treatment arms, while keeping the allocation to the control arm constant, and to two-arm trials, where RAR can be implemented asymmetrically, i.e., achieving a prespecified minimum number of allocations on the control arm to achieve sufficiently high power if the treatment effect is positive, while possibly setting a much lower minimum number of allocations on the treatment arm to be able to move the participants away from it if the treatment effect is negative.

- **Line 529-548 Comment:** The guidance document mentions that RAR designs are susceptible to inflation in Type I error probability in presence of overall time trends. This is not always the case as the control of the Type I error rate depends on many factors, such as the background response model used during the adaptation process, as well as the sample size and balanced allocation used during the burn-in stage. Using an efficient restricted randomization method at the burn-in stage can give a well-controlled type I error rate even while using RAR.

  Recommendation: We suggest to update this statement by stating what the factors are that need to be considered in the model assumption based on which RAR would be implemented that could ensure a control of the type I error rate. The guidance should emphasize including the pre-trial simulation results in the simulation report (to be accompanied with the study concept sheet) to ensure how the model assumptions affect the Type I error rates in order to select the suitable input parameters for RAR to ensure the control of false positives while implementing such a design in practice.

- **Lines 543-548. Comment:** "One approach that controls the Type I error probability is

to allow randomization ratio adaptation at only a single or small number of interim analyses, while utilizing adaptive hypothesis testing based on pre-specified weights for combining the information across trial stages. Time trends may also be addressed by using specific methodology (e.g. re-randomization tests), but an RAR design using such tests might be less powerful than a design with a fixed randomization scheme"

Recommendation: We recommend noting that when the allocation ratio changes across the study periods following one or more interim analyses, the re-randomization test should stratify by the time periods or adjust for time periods in other ways. The failure to do so is likely to result in the shift of the re-randomization distribution of the test statistics and lower the power of the test.

- **Line 549-553. Comment:** The current sentence "Given that knowledge of the RAR procedure and the adaptively selected randomization ratio could reveal information about the interim treatment effect estimate, steps should be taken to minimize what can be inferred from the adaptations (Section 3.5)." may be misleading, as it overstates the risk of information leakage and reinforces misconceptions about RAR.

  Recommendation: We suggest to rephrase the sentence, emphasizing that the exact details of the RAR adaptation and decision rules should be kept confidential similarly to other types of adaptations. In fact, changing the target allocation ratio algorithmically leads to a much smaller information leakage about the efficacy of treatments than making a permanent decision (dropping or not dropping a treatment, early stopping or not stopping a trial, etc.) by the IDMC.

- **Line 556-560. Comment:** The current sentence "Such deterministic procedures are discouraged (ICH E9) due to the high risk of bias and the potential for predicting the next treatment allocation." may be misleading, as predictability is not unique to deterministic RAR. Similar or greater predictability exists in widely used methods such as permuted block randomization.

  Recommendation: We recommend to rephrase the sentence, emphasizing that the exact details of the RAR adaptation and decision rules should be kept confidential similarly to other types of adaptations. That way, the prediction of the next treatment allocation is not notably different from (or could even be less predictable than), e.g., the current standard of the permuted block randomization (which includes predictable allocations with certainty). In fact, such deterministic adaptations are currently used without any concerns in many early phase trials, e.g., dose escalation trials, and much higher (certain) predictability exists in all open-label trials. Furthermore, there now exist methods for valid inference even for deterministic allocation procedures (e.g. Baas et al., 2025).

References

- Baas S, Jacko P, Villar SS (2025). Exact statistical analysis for response-adaptive clinical trials: A general and computationally tractable approach. Computational Statistics & Data Analysis 211(108207).
  doi: 10.1016/j.csda.2025.108207.

- International Conference on Harmonisation (1998). ICH E9: Statistical principles for clinical trials.
  url: https://database.ich.org/sites/default/files/E9_Guideline.pdf

- Kuznetsova O, Ross J, Bodden D, Cooner F, Chipman J, Jacko P, Krisam J, Luo YA, Mielke T, Robertson DS, Ryeznik Y, Villar SS, Zhao W, Sverdlov O (2025). Randomization in the age of platform trials: unexplored challenges and some potential solutions. BMC Medical Research Methodology 25:268.
  doi: 10.1186/s12874-025-02693-0.

- Kuznetsova O, Tymofyeyev Y (2012). Preserving the allocation ratio at every allocation with biased coin randomization and minimization in studies with unequal allocation. Statistics in Medicine 31(8):701-723.
  doi: 10.1002/sim.4447.

- Mukherjee A, Coad DS, Jana S (2023). Covariate-adjusted response-adaptive designs for censored survival responses. Journal of Statistical Planning and Inference 34(9):1697-1723.
  doi: 10.1177/09622802241287704

- Neyman J (1934). On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. Journal of the Royal Statistical Society 97(4): 558–625.
  doi: 10.2307/2342192.

- Pin L, Villar SS, Rosenberger WF (2024). Response-Adaptive Randomization Designs Based on Optimal Allocation Proportions. In: Chen, DG. (eds) Biostatistics in Biopharmaceutical Research and Development. Springer, Cham.

- Robertson DS, Lee KM, López-Kolkovska BC, Villar SS (2023). Response-adaptive randomization in clinical trials: from myths to practical considerations. Statistical Science 8(2):185-208.
  doi: 10.1214/22-STS865.

- Rosenberger WF, Stallard N, Ivanova A, Harper CN, Ricks ML (2001). Optimal Adaptive Designs for Binary Response Trials. Biometrics 57(3):909–913,
  doi: 10.1111/j.0006-341X.2001.00909.

- Rosenberger WF, Sverdlov O (2008). Handling Covariates in the Design of Clinical Trials, Statistical Science 223(3):404-419.
  doi:10.1214/08-STS269.

4 Operative Aspects of Trial Adaptations

General Comment

The guidance includes statements on limiting sponsor involvement in implementing adaptations and operational aspects such as IRT systems and interim database locks. However, these recommendations are inconsistent and often impractical. Adaptations typically require updates in IRT

systems for drug supply, shipment, and randomization, which generally involve sponsor participation. Complete exclusion of the sponsor conflicts with oversight obligations under ICH E6(R3). Additionally, terminology such as "formal interim database lock" and "date of sponsor unblinding" lacks clarity, and the term IVRS/IWRS should be updated to reflect current industry standards.

General recommendations

The guidance should acknowledge that full exclusion of sponsor involvement is not feasible and instead emphasize confidentiality safeguards. It should also ensure consistent terminology for IRT systems and clarify definitions for database locks and unblinding, while aligning with ICH E6(R3) oversight requirements.

Specific Comments and Recommendations

- **Lines 595-597 (Section 5.1). Comment:** It is mentioned that "the adaptations should be planned such that the sponsor can implement the IDMC recommendations regarding trial adaptations without having access to any unblinded interim results". We consider this recommendation to be very difficult to implement in practice, as adaptations will inevitably require implementation within IRT systems. These may either be with the sponsor itself, or will need to involve the sponsor for drug supply, shipment, and randomization schedule. Therefore, we appreciate some critical review of this recommendation, taking the mentioned issues in implementing such an approach into account.

  Recommendation: We suggest that the guideline should acknowledge the fact that implementing adaptations without sponsor involvement is operationally challenging. Adaptations often require updates in IRT systems for drug supply, shipment, and randomization, which typically involve sponsor participation. Instead of excluding the sponsor entirely, we propose emphasizing robust confidentiality measures (e.g., role-based access controls, use of internally unblinded functions not involved in trial activities) to prevent disclosure of unblinded data while allowing practical implementation of IDMC recommendations

- **Lines 864-866 (Section 5.6). Comment:** It is stated that "[c]linical trials with an adaptive design typically use an interactive voice or web randomization system to manage randomization and assignment of participants to treatment arms."

  Recommendation: We think it would be important to mention that these systems are in fact used in almost all global clinical trials, not only those with an adaptive design. We however acknowledge the fact that it is even more important to use an IRT system for trials with adaptive design features due to their increased complexity, not only regarding randomization but also drug supply and other critical trial elements. In addition, we would advise to use the term Interactive Response Technology (IRT) systems or IxRS (the x standing for either voice or web) instead of IVRS/IWRS, as this is more common terminology nowadays.

- **Lines 868-870 (Section 5.6). Comment:** It is mentioned here that changes in the treatment arms or randomization ratio should be done with "minimum sponsor involvement", which contradicts the statement in lines 549-550 which states that these changes should be done "without sponsor involvement". While the statement in 549-550 goes against the ICH

E6(R3) guidance that the sponsor should conduct oversight of trial-related activities, minimum sponsor involvement seems to be the only feasible approach when the sponsor is also to fulfill oversight requirements. The guideline should have a consistent position on this topic, and also mention the oversight requirements laid down in ICH E6(R3).

Recommendation: We suggest that the guideline should acknowledge the fact that implementing adaptations without sponsor involvement is operationally challenging. Adaptations often require updates in IRT systems for drug supply, shipment, and randomization, which typically involve sponsor participation. Instead of excluding the sponsor entirely, we propose emphasizing robust confidentiality measures (e.g., role-based access controls, use of internally unblinded functions not involved in trial activities) to prevent disclosure of unblinded data while allowing practical implementation of IDMC recommendations.

- **Lines 879-880 (Section 5.6). Comment:** The term "formal interim database lock" is used. It would be important to specify how a "formal interim database lock" is defined and how it is distinguished from an "informal interim database lock". ICH E9 specifically defines an "interim analysis" as "any analysis intended to compare treatment arms with respect to efficacy or safety at any time prior to formal completion of a trial" and does not distinguish "formal" or "not-formal" locks.

  Recommendation: Further clarification for the use of the term "formal" in the context of database locks would be helpful.

- **Line 941 (Section 6.2). Comment:** The term "date of sponsor unblinding" is mentioned, but depending on the needs of the trial, there may be multiple dates of sponsor unblinding varying by function.

  Recommendation: If by sponsor unblinding the term "database unblinding" is meant, which can be tied to an actual distinct date, then the latter term should be used instead.

5 Type I Error Rate Control

The following comments and recommendations aim to clarify the guidance's treatment of Type I error control and adaptive principles in exploratory and non-inferiority/equivalence trial settings, ensuring accurate interpretation and practical applicability.

Specific Comments and Recommendations

- **Lines 833-834 (Section 5.5). Comment:** The guidance states that "The principles in this guideline are also relevant in these settings to ensure reliability and interpretability of the results and subsequent decision-making based on such trials." However, the principles mentioned in this guidance are focused on controlling the statistical properties such as the type I error rate and maintaining trial integrity. Exploratory trials such as early phase dose finding trials are non-randomized and follows a deterministic procedure for dose escalation stage and even when patients are randomized to two or more doses after the dose escalation stage, the sample size is selected through clinical considerations and controlling the type I error rate is not the main focus here. Therefore, the principles of adaptive designs are quite different for such exploratory trials. The guidance document, when it speaks about

exploratory trials needs to be specific about this point when mentioning the principles of adaptive designs in such exploratory trials.

Recommendation: We recommend to clarify in Section 5.5 that maintaining the type I error rate in exploratory trials such as the dose finding trials is not the focus. We need to have enough patients to be able to assess the additional data needed to identify the optimal dose.

- **General Comment (Section 4.2):** The guidance frequently refers to the type I error rate, which is understandable. However it ignores the fact that decision errors like type I and II errors have multiple meanings in certain adaptive designs such as blinded sample size re-estimation for non-inferiority and equivalence trials.

  Recommendation: We propose to add a statement here about such nuances and the need for researchers to clarify and justify the decision errors they are controlling in such settings.

- **Line 401 (Section 4.2). Comment:** The guidance only mentions two-arm non-inferiority trial with continuous endpoints as an example, whereas this challenge of type I error rate inflation is also equally prevalent in equivalence trials also handling binomial endpoints.

  Recommendation: The example should be revised to read "(e.g., a two-arm non-inferiority or equivalence trial),"

- **Lines 139-141 (Section 4.2). Comment:** The guidance mentions that "For example, Type I error probability control requires the pre-specification of criteria for early efficacy stopping or rules for combining evidence across stages.". This is true for group sequential designs. However, Type I error rates needs to be also controlled for blinded sample size re-estimation designs in non-inferiority and equivalence trials (for example, while handling biosimilars) and do not require the pre-specification criteria for early efficacy stopping rules.

  Recommendation: Suggest adding "for group sequential designs" after "Type I error probability control".

References

- Friede T, Mitchell C, Müller-Velten G (2007) Blinded sample size reestimation in non-inferiority trials with binary endpoints. Biometrical Journal49(6):903-16.
  doi: 10.1002/bimj.200610373.

- Friede T, Kieser M (2003). Blinded sample size reassessment in non-inferiority and equivalence trials. Statistics in Medicine 22(6):995–1007.
  doi:10.1002/sim.1456.

- Mukherjee A, Coad DS, Jana S (2023). Covariate-adjusted response-adaptive designs for censored survival responses. Journal of Statistical Planning and Inference 34(9):1697-1723
  doi: 10.1177/09622802241287704.

Disclaimer

The opinions expressed by the authors (Vance W. Berger) are their own and this material should not be interpreted as representing the official viewpoint of the U.S. Department of Health and Human Services, the National Institutes of Health, or the National Cancer Institute.

The views and opinions expressed in this response letter are those of the authors and should not be interpreted as representing the official policy, position, or views of their respective institutions.

Conflict of Interest

Dr. Sofía Villar is an advisor for PhaseV, a technology company that specializes in AI algorithms to support biopharma sponsors and CROs.