EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

05 October 2026
EMA/217509/2026
Committee for Medicinal products for Human Use (CHMP)/Methodology Working Party

Reflection paper on methodological aspects in confirmatory platform trials

Draft

Draft agreed by Methodology Working Party	20 August 2026
Adopted by CHMP for release for consultation	5 October 2026
Start of public consultation	19 October 2026
End of consultation (deadline for comments)	31 January 2027

Comments should be provided using this EUSurvey form. For any technical issues, please contact the EUSurvey Support

Keywords	***Platform trials, multiplicity, non-concurrent controls, adaptive design, randomisation***

Official address Domenico Scarlattilaan 6 ● 1083 HS Amsterdam ● The Netherlands
Address for visits and deliveries Refer to www.ema.europa.eu/how-to-find-us
Send us a question Go to www.ema.europa.eu/contact **Telephone** +31 (0)88 781 6000 An agency of the European Union

© European Medicines Agency, 2026. Reproduction is authorised provided the source is acknowledged.

Reflection paper on methodological aspects in confirmatory platform trials

Table of contents

1. Purpose and Scope **3**

2. Relevant guidelines **4**

3. Key methodological topics in platform trials **4**

3.1. Pre-specification 4

3.2. Randomisation 5

3.2.1. Allocation ratios and probabilities 5

3.2.2. Multi-stage randomisation 6

3.2.3. Re-randomisation of participants 6

3.3. Controls 7

3.3.1. Common controls 7

3.3.2. Concurrent versus non-concurrent controls 8

3.3.3. Change in control treatment 8

3.4. Blinding 9

3.4.1. Unblinding during study conduct 10

3.5. Type I error control 10

3.5.1. Requirements for strong FWER control 11

3.5.2. Risk of multiple false-positive decisions 11

3.5.3. When confirmatory hypotheses can be separated into unrelated families 11

3.5.4. Pre-specification of testing strategy and objectives 12

3.6. Trial operating characteristics 12

3.6.1. Power and sample size 12

3.6.2. Further operating characteristics 12

3.7. Estimands 13

4. Complexity from Multiple Features **13**

1. Purpose and Scope

Platform trials have become more common over recent years and are likely to play an increasingly important role in marketing authorisation applications. Design elements, which frequently occur in platform trials, such as adaptive designs and common controls increase the complexity in planning, conduct and reporting of clinical trials. This makes dedicated guidance necessary to enable adequate planning and assessment of platform trials such that they can be used as confirmatory trials. In this reflection paper aspects specific to or highly prevalent in platform trials are highlighted. Their impact on the reliability of the results in terms of bias, type I error, and internal, as well as external, validity is discussed. These aspects include the choice and definition of controls, randomisation, blinding and adaptive design elements. The aim of this reflection paper is to complement existing guidance documents, e.g. on multiplicity and adaptive designs, not to replace or revise them.

For the purpose of this reflection paper, we define "platform trial" as the generic term for a multi-arm clinical trial, which is executed as a **single clinical trial but may address more than one primary objective** with the following key elements:

- **Multiple treatments** are investigated within one trial in separate experimental arms in a **single target population**.
- **Randomisation** is used for treatment allocation. Randomisation probabilities may change over time, e.g. due to a changing number of experimental arms.
- Treatments are compared to a control. This may be implemented through either a comparison to a **common, shared control arm** or to **separate control arms for each treatment**.
- **Arms may be added or terminated** at different times, and there can be multiple analysis time points. A multi-arm trial with all arms starting and ending at the same time can be considered a platform trial as well.

In a platform trial, all objectives are preferably specified under a single protocol, which may sometimes be referred to as a master protocol. However, the definition of a master protocol is wider than that of platform trial in this reflection paper. It may encompass multiple protocols, each corresponding to a sub-study, which are grouped under a single protocol for administrative and/or logistical purposes, that could potentially stand alone as separate studies. In this reflection paper the term protocol will be used throughout the document regardless of how the protocol is structured (i.e., if it is specified in one or multiple documents). For these reasons, the term master protocol with the more general definition is not used in this reflection paper. These operational aspects and aspects related to clinical trial applications are not of relevance for the methodological considerations but are discussed elsewhere (see Section 2.).

The aim of this reflection paper is to describe under which concrete circumstances and methodological constraints platform trials can provide reliable evidence for regulatory decision-making. The primary focus of this reflection paper is on **confirmatory trials**, i.e., trials that are designed to provide pivotal evidence for regulatory decision making. Other trials with similar design features and exploratory platform trials are not explicitly in scope, although they may share the same key characteristics and hence benefit from these considerations as well. Key methodological topics and design features of relevance in platform trials to guide study planning, protocol development, conduct of study and analyses of results are discussed, with the aim to obtain reliable results to support regulatory decision making.

2. Relevant guidelines

All general guiding principles for clinical trials naturally apply to platform trials as well, and should be considered during planning, conduct and analysis of clinical trials. More specifically, this reflection paper needs to be read in conjunction with the following methodological regulatory guidelines:

- ICH E9: Statistical principles for clinical trials (CPMP/ICH/363/96)
- ICH E10: Choice of Control Group in Clinical Trials (CPMP/ICH/364/96)
- ICH E20: Adaptive designs for clinical trials (EMA/CHMP/ICH/206586/2025)
- Reflection paper on methodological issues in confirmatory clinical trials planned with an adaptive design (CHMP/EWP/2459/02)
- Multiplicity issues in clinical trials (EMA/CHMP/44762/2017)
- Guideline on data monitoring committees (EMEA/CHMP/EWP/5872/03 Corr)
- Questions and answers on data monitoring committees' issues (EMA/CHMP/470185/2020)

Operational or clinical trial application aspects will not be covered in this reflection paper. Please refer to the Complex clinical trials Q&A (EMA/298712/2022) and the following guidance:

- Recommendation Paper on the Initiation and Conduct of Complex Clinical Trials (hma.eu)
- Frequently asked questions on Complex Clinical Trials and CTIS (hma.eu)

3. Key methodological topics in platform trials

3.1. *Pre-specification*

Since platform trials are complex to design and conduct, additional efforts are required at the planning stage to consider all relevant aspects. It is crucial that the objectives, estimands and the overarching testing strategy in the platform trial are thoroughly documented at the start of the study as far as possible. This documentation is essential to confirm that the design corresponds to the objectives and to determine over which hypotheses the type I error has to be controlled. If new arms with potentially new estimands, new specific analyses strategies or new objectives are foreseen to enter a platform trial over time, these should be described and justified as early and as thoroughly as possible. Adherence to these pre-specifications is of particular importance in a highly complex and adaptive trial design. Modifications at a later point in time might potentially be data-driven. It usually is very hard or even impossible to prove the opposite. Lack of pre-specification could render a confirmatory platform trial as exploratory only.

Examples of aspects that are specific to platform trials, and require clarity and justification, include:

- The definition of all primary objectives and primary estimands;
- The subset of the total study data that will be utilized to inform each objective;
- The strategy for adapting the trial design based on emerging data, including criteria for adding or removing arms;
- The approach to handling the multiple primary objectives and controlling type I error rates in the platform trial;
- The timing of analyses (interim and primary analyses) for all objectives.

Randomisation is a key element of statistical methods in clinical trials. It facilitates valid causal inference and indirectly accounts for known and unknown prognostic factors. Together with blinding it helps to conceal treatment allocation. Hence, randomisation is vital to avoid or reduce bias and forms a basis for valid statistical inference and confirmatory analyses (see ICH E9, ICH E10). ICH E9 states that the unique pre-planned assignment of treatment, or treatment sequence, to subject should be clear, and that the randomisation schedule should be reproducible. In platform trials, some aspects of randomisation are more complex, e.g., due to potential changes in the number of experimental arms. Hence, it may not be possible to fully pre-specify the randomisation plan in advance. Nevertheless, the principles of the randomisation procedure should be specified before study start and all details carefully documented during study conduct. The high-level principles essential for understanding the approach should be included in the protocol and the specific details documented in a "randomisation plan" or equivalent. In general, circumstances under which the clinical trial is conducted evolve over time. This potentially affects e.g. the clinical experience with a specific treatment, and the population enrolled to a clinical trial. This is true for any trial, but particularly relevant in a platform trial context which typically has an increased duration. The trial itself can also be altered, e.g., through a change of the enrolling experimental arms during the trial, changing allocation ratios or even changing the control arm, which may complicate the interpretation.

The population enrolled in the trial might be impacted by a participant's willingness to participate depending on changes in allocation probabilities, such as an increased likelihood of being randomised to active treatment, or on changes in available treatments, for instance by the introduction of a particularly promising treatment or a medicinal product with an expected high risk of side effects. Thus, selection bias may be induced and may change over time. Randomisation may help to overcome the induced issues of these fluctuations to some extent if only concurrently generated groups of participants are compared (see Section 3.3.2.). This is also pointed out in ICH E10 in general, and the Complex clinical trials Q&A (EMA/298712/2022) in specific, "randomised controlled clinical trials with an internal concurrent control arm are considered the standard in trial design where appropriate and feasible". While randomisation addresses the issue of causal interpretation and isolation of a treatment effect, it is worth noting that changes in circumstances within a clinical trial over time may complicate and even compromise the interpretability of the treatment effect.

3.2. *Randomisation*

3.2.1. Allocation ratios and probabilities

Changes in allocation probabilities can occur for several reasons but in platform trials this issue arises specifically due to the addition and removal of arms. In a platform trial, the relevant comparison is typically between an experimental arm and a control arm within the trial. Thus, the **allocation ratio between these arms within the trial** is central to the inference. Allocation ratios should ideally be held constant for all comparisons of interest, even when the allocation probabilities are changed by adding or dropping experimental arms. If the allocation ratio is constant for a comparison of interest, standard inference applies. Equal allocation to all arms, i.e., $1{:}\ldots{:}1$ allocation, is a common choice to achieve this. Complexity is added if the study design causes the allocation ratios to change over time. The choice of allocation ratio may also depend on other factors, such as the need for sufficient safety data for the medicinal products under investigation. Depending on the trial design and purpose, more efficient allocation ratios than equal allocation exist. The square root allocation first introduced by Dunnett is one example of an allocation rule for multiple comparisons to a single control. It is given by allocating subjects with a ratio of $1{:}\sqrt{k}$ to each experimental arm and the control arm, where k is the number of experimental arms, resulting in an allocation of $1{:}\ldots{:}1{:}\sqrt{k}$). This reduces the correlation of test statistics and increases power. While this allocation rule might be beneficial in platform trials with

a fixed number of arms and without any adaptations over time, this is not always the case in adaptive platform trials, as it is desirable to keep allocation ratios constant over time for each comparison of interest.

Not only the allocation ratio but also the choice of the adequate randomisation scheme requires special attention in platform trials. For instance, block randomisation, which is the most common randomisation method, might cause challenges. With multiple arms the block size may become unpractically large, especially with a small number of participants. This means that the block randomisation becomes a simple, unconstrained randomisation approach and the desired allocation balance between arms might suffer to an unknown extent. Furthermore, when square root allocation is used, standard block randomisation may not even be feasible. Another example is the use of dynamic randomization methods such as minimisation. In this case, the addition or removal of arms adds further complexity to an already complex method and may have consequences for statistical inference. Hence, such approaches are usually discouraged in platform trials. Operational implementation of the randomisation scheme must be thoroughly considered.

Changes in allocation ratios over time can in theory be addressed in the statistical analysis, for example using an analysis stratified by time periods, which are defined by the different randomisation ratios. However, some design features may add too much complexity, making it difficult or impossible to properly adjust for changes in the allocation ratio in the analysis. For this reason, it is recommended to keep the allocation ratios for each comparison of interest fixed. Consequently, (response) adaptive randomisation (RAR) is not supported.

3.2.2. Multi-stage randomisation

Randomisation might also be conducted in two (or more) stages. In the first stage, participants are randomised to a group of treatments, e.g. based on route of administration. In the second stage, participants are randomised to either the treatment (or one of the treatments) or control in that group. The rationale for multi-stage randomisation is of relevance to understand its implications to study conduct and the statistical analysis and should be included in the protocol.

Where all participants are eligible for all treatments, and two-stage randomisation is applied solely for operational purposes (e.g. to facilitate partial blinding), the resulting allocation process is otherwise equivalent to a standard randomisation approach. In that case, concurrently enrolled controls might be pooled, if all aspects of study conduct are identical for all participants, in aspects key to the objectives. However, differences may occur even if the study conduct is the same. For example, the expectations of participants may differ for different groups of treatments, thus, potentially leading to different placebo effects or different dropout patterns. An upfront discussion of these aspects is needed together with a plan to assess the similarity of randomisation groups if pooling across different groups is planned. The latter might be difficult due to the small sample size in each of the control arms. If other reasons apply, e.g. as inclusion or exclusion criteria differ for different treatment options, the control arm of one treatment may not be representative in a comparison to another treatment and control arms cannot be pooled in most scenarios. See further discussion below in Section 3.3.1. .

3.2.3. Re-randomisation of participants

Due to the possible longer duration, the multiple treatment options and the dynamic nature of a platform trial, the possibility to re-randomise participants might be considered by the sponsor. Re-randomisation adds further complexity to the trial. It usually changes the questions that can be answered (and hence impacts the indication). It may lead to situations where the new target population is difficult to define, as re-randomised participants had previously been treated with non-

marketed, experimental treatments. It might further induce statistical dependency issues and bias in the estimated treatment effect for the target population. Hence, it is usually discouraged in a confirmatory setting, and if planned to be used, it should be well justified, also on methodological grounds, before being implemented and would need to be pre-specified in the protocol before study start. Furthermore, the process related to the re-randomisation in terms of conditions for when it might be applicable as well as methodological considerations should be pre-specified and well described in the protocol.

3.3. *Controls*

The choice of a control arm is of fundamental importance as it defines which questions can be answered. Guidelines relevant to choice of control arm such as ICH E10 apply. A platform trial may provide answers to multiple questions concerning different treatments and, hence, the definition of controls is more complex in platform trials than in standard two-arm randomised controlled trials, given that multiple experimental arms are included. This leads to different options to define the control arm for a specific comparison:

1) use a separate, independent control arm for each of the experimental arms that address a specific comparative objective,

2) use a common control arm for all experimental arms together, or

3) use combinations or variants of these two options.

Another distinct aspect in platform trials is that randomisation is typically open only to certain arms at a given point in time. In particular, enrolment into a control arm may be open for longer than enrolment into the respective experimental arm. Concurrent and non-concurrent controls are discussed in Section 3.3.2. .

3.3.1. Common controls

One of the most relevant and typical features of platform trials is the use of a common control arm, which is also referred to as a shared control arm. Rather than randomising subjects to either an active treatment or a dedicated control treatment (or placebo) on a per treatment basis, one randomly allocates subjects to any of the enrolling experimental arms or to the shared control arm. This may allow to achieve the same power to detect a specific treatment effect (marginal power) with less subjects overall compared to separate controls for each treatment (see Section 3.4.) and to treat fewer subjects with the control treatment. However, the use of a common control arm will also lead to correlation between the test statistics of the different comparisons, which increases the probability of drawing simultaneous multiple false conclusions, while leading to a lower probability of at least one false-positive rejection within the trial. This is discussed in Section 3.5.2. .

A sometime proposed approach is to randomise to a matching control treatment per experimental arm and to pool the control arms for the primary analyses. Independent controls may be justified through operational aspects as discussed in Section 3.2.2. , and if they are planned to be pooled, it is paramount to discuss potential differences induced by different control treatments already at the planning stage. In general, any differences in study conduct, such as the schedule of assessments, are usually issues that prevent pooling.

3.3.2. Concurrent versus non-concurrent controls

Concurrent controls are here defined as subjects in the control arm who have a positive probability to be randomised to the experimental arm of interest at the time of randomisation. In practice, this means that participants were all randomised at the same time and were eligible to be randomised to the arms with respect to inclusion/exclusion criteria and participants provided informed consent for all arms. A strong preference is given to concurrent controls as this enables that a causal conclusion can be achieved without additional assumptions. Another advantage is that it may reduce the correlation between test statistics, as it limits the overlap of the respective participants in the control arm used to calculate the test statistics. If the trial is blinded, it might additionally help to mitigate practical issues with blinding as discussed in Section 3.4. .

It is acknowledged, though, that unique situations exist where it might be acceptable to increase the power by using non-concurrent controls. This needs to be pre-planned and well justified in the trial protocol. Methods to reduce the temporal bias should be pre-specified as early as possible (see Section 3.1) together with supplementary and sensitivity analyses to assess the impact of the underlying assumptions. These additional analyses should include other adjustment methods, and analyses solely based on concurrent controls. Given the strong reliance on untestable assumptions, non-concurrent controls are usually not acceptable for the primary analysis of the primary endpoint and are more suitable for supplementary analyses or potentially for secondary endpoints. This approach might be especially relevant for trials with primary endpoints which can be shown with sufficient power based on concurrent controls, but where important secondary endpoints exist that are rarely observed.

External controls, i.e. data from other sources than the platform trial, are not specifically discussed in this reflection paper. While the most reliable approach is to use concurrent controls, internal, non-concurrent controls in platform trials would be preferred compared to external controls. The reason is that non-concurrent controls within a platform trial were enrolled under the same protocol and hence share many commonalities. Hence, assumptions required to augment concurrent controls and to enable valid inference may be more plausible when trial conditions are governed by a common protocol, leading in general to less bias. Temporal drifts or time bias might exist but attempts to model the temporal changes are possible using other experimental arms, if overlapping with both concurrent and non-concurrent controls.

3.3.3. Change in control treatment

In platform trials it is sometimes planned to change a treatment which met its primary endpoint within the platform trial to become the new control treatment. The presented rationale for this adaption is that the standard of care and, hence the scientific question of interest, has changed. While this may in principle be understandable, it comes with various issues.

First, the fact that efficacy was shown for a new treatment does not mean that a positive benefit-risk profile of the new treatment has been established and that the product will eventually be authorised. Benefit-risk decisions require a careful analysis and assessment process. Only after a treatment has been authorised, one could consider it as the new control, in order to not expose the trial participants to unnecessary risk, but also not to hamper the clinical interpretation of the results. Exceptions from this rule may apply in exceptional circumstances.

Second, a change in control treatment introduces substantial statistical issues and might compromise the interpretation of trial results. The primary analysis and related operating characteristics, including the sample size, for all arms currently ongoing in the trial may require modification. Furthermore, it is not evident whether the initial control arm would be combined with the new control arm, or whether

they would be assessed separately. The former is methodologically difficult and hampers the interpretability of the derived results, since the effect before and after the modification must be taken into account. It is also very difficult to pre-specify such changes and to control the type I error. The latter approach leads to power issues and may not be considered adequate as not all enrolled participants are used (equally) to support the objectives and decision making.

Hence, the control arm should be changed only in very exceptional circumstances. This principle applies universally to all changes to the control regardless of trial phase. In exploratory trials methodological issues might be less critical. However, it is likely to be questionable whether an unauthorized treatment is safe and efficacious when it only met the trial objectives in an exploratory trial and whether it could hence serve as a new control treatment. Therefore, it may be even less appropriate to promote an experimental treatment to become the new control treatment in exploratory trials.

Hence, for all the aforementioned reasons, it is discouraged to change the control treatment for arms that have already started enrolling. If it cannot be avoided to change the control arm, usually a new trial would be the preferable option. However, it may be justified to define a different control arm from what was initially planned in the protocol for newly planned experimental arms which have not yet started, if the standard of care for that specific indication has changed during the ongoing clinical trial.

3.4. *Blinding*

If the trial includes multiple test groups that are to be compared to a common control group, and blinding is to be achieved for all comparisons, this requires that the subjects are blinded with respect to all the groups that they could have been randomised to. While this is in theory possible by performing a multiple-dummy type of blinding procedure, where participants receive placebo treatment for all experimental arms that they do not belong to, such a procedure may be too burdensome for the participants.

If experimental arms start at different time points this adds additional complexity to the blinding procedure. Since participants have the right to know the potential treatments they may be randomised to, blinding is only possible against arms that the participant is eligible for at the time of randomisation.

A compromise may be to partially blind the trial. For example, participants will know their allocation to a specific route of administration or even a specific treatment/control comparison but remain blinded regarding their allocation to treatment or the corresponding control. They will know, however, that they were not allocated to another route of administration, or another treatment and the corresponding control treatment. Operationally, this can be achieved with a 2-stage randomisation process (see also Section 3.2.2.). Pooling of concurrently enrolled control arms may then be possible upon justification (see Section 3.3.1.).

Open-label trials are prone to operational bias. Hence, if the trial is conducted open label, it is important to ensure that the protocol is adhered to. As platform trials often include adaptive features and go through design changes while the trial is ongoing, it is particularly important to ensure that these are not motivated by information stemming from the open-label nature. Regardless of whether participants and local investigators are blinded, measures to blind the central study team should be pre-specified in the protocol and/or a dedicated blinding plan. Clear rules for data access and decision making should be in place. An independent data monitoring committee (DMC) is generally recommended but might not be sufficient to solve all issues.

3.4.1. Unblinding during study conduct

In platform trials, the primary analyses for the various primary objectives may be conducted at different points in time, which may introduce bias and compromise the integrity of the study. The main reason for this is partial or full unblinding of the shared control arm results and potentially also of the results of other experimental arms. If subsequent adaptations of the trial use this information, either implicitly or explicitly, without adequate pre-specification, this could lead to a type I error inflation and biased results. Hence, great care should be taken to maintain blinding as much and as long as possible and to fully pre-specify the study conduct as far as possible.

When data is planned to be analysed at the participant level and on aggregated level during the trial, it is recommended to plan the trial such that as many evaluations and decisions as possible remain blinded to participants, site personnel and sponsor. It will usually be expected that an independent DMC is implemented (see Guideline on data monitoring committees, EMEA/CHMP/EWP/5872/03 Corr, and Questions and answers on data monitoring committees issues, EMA/CHMP/470185/2020), an adequate firewall to shield the decision makers in the trial team is in place, and that maintaining full blinding until all concurrently recruiting experimental arms are completed is considered. Having different analysis teams or even different study teams for different interventions studied in the platform trial might further help to limit data leakage possibilities. Importantly, it should be acknowledged that maintaining the study integrity is not always possible. It is not a realistic expectation that results are not known if parts of the trial are submitted for Marketing Authorisation or published elsewhere when other parts are still ongoing. This may compromise robustness of the results and impact their reliability for regulatory decision making.

Note that in platform trials with time-to-event endpoints and event-driven analyses additional care is needed. In this case, even the availability of standard monitoring information such as the pooled number of events for one specific intervention and its concurrent control might lead to unblinding and data leakage, as events could be attributed to each arm if multiple arms run concurrently. In that case, it might be preferable to blind the study teams to the accrual of events and only report the projected analysis times to the teams, while tasking external parties with the actual monitoring of the event accrual and updates of the projected analysis times.

The above-mentioned measures may help to avoid or at least reduce the leakage of data. If data leakage cannot be fully excluded, further measures might be required to reduce the impact of data leakage on the conduct of study and the robustness and credibility of results. For example, one could consider requiring that analyses are conducted at the same time for all concurrently running arms, which essentially means possibly postponing the analysis time point.

Furthermore, any protocol amendments, in particular after some results have already become available, require a strong justification in order not to compromise the integrity of the ongoing parts of the platform trial.

Some of these sources of bias and risks to study integrity cannot be fully mitigated, as they are integral to the motivation to conduct a platform trial. Therefore, such designs may only be feasible in settings where objective endpoints can be used, and good study conduct and participant adherence can be ascertained.

3.5. *Type I error control*

This section aims to determine under which conditions Family-Wise Error Rate (FWER) control across the hypotheses related to all primary objectives may or may not be necessary in platform trials to support confirmatory claims. Additional points of attention with multi-arm trials performed to generate

evidence in support of more than one regulatory decision, such as the increased risk of multiple false-positive decisions, i.e., wrongly rejecting multiple null hypotheses, are also discussed.

3.5.1. Requirements for strong FWER control

The consequences of multiplicity in platform trials are determined by the overall question(s) the trial is intended to answer and the (regulatory) decisions it may lead to and requires a prospective, multi-disciplinary justification. In platform trials strong FWER control (see Guideline on multiplicity issues in clinical trials; EMA/CHMP/44762/2017) is not required by default across all hypotheses. However, strong FWER control is required across hypotheses that are related to the same overall question and the corresponding conclusions contributing to a regulatory decision.

3.5.2. Risk of multiple false-positive decisions

In every instance where there is more than one comparison, one carries the risk of drawing multiple erroneous decisions (simultaneously). It has been shown that trials comparing multiple treatments to a common control reduces the risk of at least one erroneous conclusion, as measured by the FWER, compared to a platform trial with individual control arms for each experimental arm (when in both scenarios no multiplicity adjustment is used). The latter trial scenario has the same error distribution as separate independent trials that compare each treatment to an individual control group. From the perspective of FWER control, a platform trial with a common control arm hence seems preferable. However, using a common control increases the probability of drawing multiple erroneous conclusions at the same time, as measured by the k-FWER with $k \geq 2$. This is due to the correlation between the test statistics. The expected number of false conclusions per family remains the same.

Controlling this potential increase in the risk of multiple false-positive errors is technically possible by imposing stricter limits on the individual significance level associated with each experimental arm. However, doing so would require stringent thresholds which may be more conservative than a Bonferroni correction. An alternative to reduce the risk of multiple false-positive errors is to increase the sample size in the control arm. A second confirmatory trial in the same or a similar setting, potentially with different objectives, will usually be the best option to counter the issue of (multiple) false-positive errors and bias. Additionally, it improves the overall strength of evidence when the trials relate to the same treatment.

3.5.3. When confirmatory hypotheses can be separated into unrelated families

Two conclusions from comparative hypothesis tests contributing to regulatory decisions are here considered unrelated if the outcome of one comparison should have no influence on the outcome of the other; correspondingly, hypotheses are considered unrelated if the outcome of the statistical test regarding one hypothesis has no impact on the evaluation, interpretation, or regulatory decision associated with the other. In other words, knowledge of the outcome of one hypothesis test does not directly influence the conclusion, and related regulatory decision of the other one.

The most straightforward examples are as follows:

- If the two experimental arms in a pivotal trial are different doses of the same medicine, with the aim of choosing at least one dose to be included in the label, the hypotheses are related.
- If the two experimental arms are two new active substances compared against a control arm and the decision to license one of them is made independent of the results of the other one, the hypotheses concerning the efficacy of either active substance are unrelated.

Unlike statistical independence, the relatedness of hypotheses is not a mathematical property. Unrelatedness, therefore, is based on the clinical context and the overall rationale of the clinical trial, which is reflected in the design, objectives and experimental arms and hypotheses of interest. The relatedness of hypotheses essentially refers to the question of whether the hypotheses refer to a common "overarching hypothesis" that is reflected in the corresponding regulatory decision. In most scenarios, hypotheses concerning the same active substance are related, because in most situations the relevant regulatory decision will be at the active substance level.

If the hypotheses tested in a trial can be separated into families of unrelated hypotheses, the control of the type I error rate may not necessarily be required across families. Yet, for hypotheses within a given family the usual criterion, i.e., the need for strong FWER control, applies.

3.5.4. Pre-specification of testing strategy and objectives

As discussed in Section 3.1, the overarching objectives, overarching testing strategy, and a plan for how the results will be used, need to be clearly pre-specified and described in the protocol. This pre-specification is essential and must be justified in relation to the corresponding regulatory decisions they intend to inform. If any such information is missing in the protocol, it is usually not possible to determine if the hypotheses can be separated into distinct 'families' of unrelated hypotheses. Generally, this omission implies that FWER control will be necessary.

3.6. *Trial operating characteristics*

3.6.1. Power and sample size

The significance level depends on the specifics of the platform trial and naturally impacts the power and sample size. In addition to that, power and sample size require additional attention in a platform trial as there is more than one type of power. The marginal power is the power to reject a *specific* hypothesis test. This is the standard power definition as used in other trials as well. In addition, the disjunctive power is defined as the probability to reject *at least one* false null hypothesis, and the conjunctive power is defined as the probability to reject *all* false null hypotheses. The relevance of each of these definitions depends on the objectives of the trial. Thus, power and sample size always need to be justified based on the relevant power definition(s).

The use of common controls impacts the power and sample size beyond the impact of the significance level. For a specific targeted conjunctive power usually a smaller overall sample size is needed in a platform trial with shared controls than in a trial with independent controls (or in separate trials). The same is true for the marginal power due to the common control. To reach a specific disjunctive power, however, a platform trial with shared controls usually needs a larger sample size than with individual controls. This needs to be considered at the design stage as well.

Hence, a sample size justification for a platform trial should explicitly define which power is targeted and whether the correlation between the test statistics was considered. The overall power and sample size justification should be aligned with the overall objectives and multiplicity strategy as discussed in Section 3.5. .

3.6.2. Further operating characteristics

In addition to the type I error rate (Section 3.5.) and the different possible definitions of power (Section 3.6.1.), further operating characteristics should be considered in a platform trial. These include plausible platform trial trajectories, including arm entries and discontinuations, number of arm

open for recruitment at any given time point, recruitment rates over time, the impact of shared controls (Section 3.3.1.), the expected sample size per arm, the expected overall sample size, the expected duration of each arm, and the expected overall trial duration. It is important to fully characterize the relevant operating characteristics in the protocol to allow for comprehensive regulatory assessment. For this purpose, simulations may be required and should be included in the protocol as appropriate.

3.7. *Estimands*

The concept of estimands, as formalised in the International Council for Harmonisation E9(R1) guideline, is essential for ensuring clarity in the definition of treatment effects in clinical trials, including complex designs such as platform trials. The estimand definition(s) should correspond to and be consistent with the overarching study objectives and hypothesis testing strategy outlined in the protocol. By explicitly specifying the population, treatment (at least, intervention and comparator), variable, and handling of intercurrent events, the estimand framework provides a structured approach to align trial objectives with the scientific questions of interest. This is particularly important in platform trials, because there are multiple primary objectives and, hence multiple estimands, and it may need to be carefully considered for each estimand if it can be appropriately assessed with a study design where, e.g., experimental arms may be added or dropped, or the target population might evolve. Within a platform trial, each experimental arm is typically associated with its own estimand, reflecting a specific clinical objective; while these estimands may share common components, such as the target population, they may differ in other attributes depending on the scientific question. There may be differences in intercurrent events between the arms, for example due to fundamental differences in treatment regimens. These require particularly careful description and justification. Ensuring transparency across these related estimands is critical. Furthermore, in the rare exemptions where the control arm changes over time (for example, due to the emergence of a new standard of care), careful consideration is needed to preserve interpretability: estimands may need to be prospectively defined in a way that is robust to such changes (e.g., referencing a class of control treatments), or alternatively, new or revised estimands may be introduced to reflect the updated comparison. In all cases, clear documentation of how estimands evolve alongside the trial is essential to maintain the validity and interpretability of the trial results.

4. Complexity from Multiple Features

While platform trials may be suitable for regulatory decision making, it is important to ensure that their results are sufficiently robust and can be unambiguously interpreted for the specific treatment that is subject of decision making. Trials incorporating more complex design features and operational requirements are inherently more challenging to assess.

Therefore, the assessment whether a platform trial provides confirmatory evidence is based on a holistic evaluation of all the methodological and operational aspects and their interdependencies. In other words, while design or operational features may be justifiable and acceptable in isolation, their complex interplay contributes to increase uncertainty and is, hence, of relevance and requires specific attention. For instance, design modifications while the trial is ongoing may make pre-specification of the overall objective challenging and consequently, corresponding multiplicity and sample size considerations difficult. Or as an alternative example, multiple primary analyses at different points in time in an (open-label) platform trial may compromise the study integrity and introduce substantial bias. Depending on the circumstances, the combination of the design elements may render a clinical trial initially planned as confirmatory an exploratory trial. Simpler designs are preferred if no substantial advantage can be expected with a complex design. Therefore, all relevant aspects should

be discussed, addressed and justified at the planning stage to allow for a comprehensive regulatory assessment.