Early phase clinical trials extension to the guidelines for the content of statistical analysis plans

An Early Phase Clinical Trials Extension to the Guidelines for the Content of Statistical Analysis Plans

Journal:	*BMJ*
Manuscript ID	BMJ-2021-068177.R1
Article Type:	Research methods and reporting
Date Submitted by the Author:	13-Dec-2021

Complete List of Authors:	Homer, Victoria; University of Birmingham, Cancer Research (UK) Clinical Trials Unit Yap, Christina; The Institute of Cancer Research, Clinical Trials and Statistics Unit Bond, Simon; Cambridge Clinical Trials Unit Holmes, Jane; University of Oxford, Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences Stocken, Deborah; University of Leeds , Leeds Institute of Clinical Trials Research Walker, Katrina; University of Leeds Leeds Institute of Clinical Trials Research, Clinical Trials Research Unit Robinson, Emily; Royal Marsden NHS Foundation Trust, Royal Marsden Clinical Trials Unit Wheeler, Graham; Imperial College London, Imperial Clinical Trials Unit Brown, Sarah; University of Leeds Leeds Institute of Clinical Trials Research, Clinical Trials Research Unit Hinsley, Samantha; University of Glasgow, Cancer Research UK Glasgow Clinical Trials Unit Schipper, Matthew; University of Michigan, Departments of Radiation Oncology and Biostatistics Weir, Christopher; University of Edinburgh, Edinburgh Clinical Trials Unit Rantell, Khadija; Medicines and Healthcare Products Regulatory Agency Prior, Thomas; Janssen Research and Development LLC, Early Development Oncology Statistics Department Yu, Ly-Mee; University of Oxford, Department of Primary Care Health Sciences Kirkpatrick, John; Roche Products Ltd Bedding, Alun; Roche Products Ltd Gamble, Carrol; University of Liverpool, Biostatistics Gaunt, Piers; Univeristy of Birmingham, Cancer Research UK Clinical Trials Unit

Keywords:	Clinical trial, Early Phase, Statistical Analysis Plan, Guidelines

An Early Phase Clinical Trials Extension to the Guidelines for the Content of Statistical Analysis Plans

Victoria Homer, Biostatistician, Cancer Research Clinical Trials Unit, University of Birmingham

Christina Yap, Professor of Clinical Trials Biostatistics, Clinical Trials and Statistics Unit, The Institute for Cancer Research, London

Simon Bond, Lead Senior Statistician, Cambridge Clinical Trials Unit, Cambridge

Jane Holmes, Senior Medical Statistician, Oxford Clinical Trials Research Unit, Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford

Deborah Stocken, Professor of Clinical Trials Research, Clinical Trials Research Unit, Leeds Institute of Clinical Trials Research, University of Leeds

Katrina Walker, Senior Medical Statistician, Clinical Trials Research Unit, Leeds Institute of Clinical Trials Research, University of Leeds

Emily J. Robinson, Lead statistician, Royal Marsden Clinical Trials Unit, The Royal Marsden NHS Foundation Trust

Graham Wheeler, Senior Lecturer in Clinical Trials Statistics, Imperial Clinical Trials Unit, Imperial College London

Sarah Brown, Principal Statistician, Clinical Trials Research Unit, Leeds Institute of Clinical Trials Research, University of Leeds

Samantha Hinsley, Lead Biostatistician, Cancer Research UK Glasgow Clinical Trials Unit, University of Glasgow

Matthew Schipper, Research Professor of Biostatistics, Departments of Radiation Oncology and Biostatistics, University of Michigan

Christopher J. Weir, Professor of Medical statistics and Clinical Trials, Edinburgh Clinical Trials Unit, Usher Institute, University of Edinburgh

Khadija Rantell, Statistical Advisor, Medicines and Healthcare products Regulatory Agency

Thomas Prior, Principal Statistician, Early Development Oncology Statistics Department, Janssen Research & Development

Ly-Mee Yu, Associate Professor, Primary Care Clinical Trials Unit, University of Oxford

John Kirkpatrick, Senior Principal Statistician, Roche Products Ltd

Alun Bedding, Associate Director, Roche Products Ltd

Carrol Gamble, Professor of Medical Statistics, Liverpool Clinical Trials Centre, University of Liverpool

Piers Gaunt, Principal Biostatistician, Cancer Research Clinical Trials Unit, University of Birmingham, Vincent Drive, Edgbaston, Birmingham, B15 2TT, p.gaunt@bham.ac.uk, ORCID ID: 0000-0001-8665-300X

Key Points

Standfirst:

This paper reports guidelines for the content of Statistical Analysis Plans (SAPs) for early phase clinical trials, ensuring specification of the minimum reporting analysis requirements, by detailing extensions (11 new items) and modifications (25 items) to existing SAP guidance by Gamble et al. following a multi-stakeholder review.

Word count: 47/100

Summary points:

Guidance for the content of SAPs for clinical trials was published in 2017 and focused on late phase, randomised controlled trials.

The existing guidelines have been extended to broaden their applicability to early phase (phase I and non-randomised phase II) clinical trials.

This extension is based on: existing guidance; a comprehensive search to identify existing published protocols, SAPs and SAP guidance; a survey of clinical trial funders and regulators; a survey of current practice of statisticians within UKCRC registered CTUs; a critical appraisal and expert review meeting; and a pilot of the proposed guidelines.

Of the 55 items originally stated in the current SAP content guidance: 30 have remained unchanged; 25 have been modified; and a further 11 new items have been proposed to ensure comprehensive and appropriate guidance for early phase clinical trials.

Word count: 134

Background and Scope of Early Phase SAP Guidance Extension

This paper details guidelines for the content of Statistical Analysis Plans (SAPs) for early phase clinical trials, presenting an extension to "Guidelines for the Content of Statistical Analysis Plans in Clinical Trials" by Gamble *et al.*(1)

Early phase (phase I and non-randomised phase II) clinical trials aim to determine the safety and initial indicators of efficacy of interventions prior to conducting potentially practice-changing phase III clinical trials. The undertaking of definitive late phase clinical trials is often a lengthy and costly process since these clinical trials ensure full-scale evaluation of the interventions and may also involve cost-effectiveness analyses. Definitive clinical trials are predicated on accurate and robust conclusions from early phase clinical trials, with flaws in design and analysis potentially a reason for interventions failing to demonstrate a benefit in phase III clinical trials. Consequently, the design, conduct, and analysis of early phase clinical trials does not solely impact that specific study. Conclusions from early phase clinical trials have implications for all related subsequent clinical trials, as such these studies must be performed to the highest standards of rigour and quality, to ensure correct decisions are taken forward.

Historically, phase I clinical trials were conducted without significant statistical involvement and conformed to rule-based designs, for example, the 3+3 design, to determine the maximum tolerated dose (2,3). Recent recommendations propose that phase I studies should employ model-based designs (4), such as the continual reassessment method (CRM) (5–8), or model assisted designs, such as a modified toxicity probability interval (mTPI) design (9). In addition, randomised dose finding phase I clinical trials (such as those which randomise to attain the optimal doses or dose schedules once safety has been assured (10)) and single arm phase II designs (11) are being used, all of which require significant statistical input before, during and at the analysis stage of the clinical trial. The use of these more statistically involved clinical trial designs has been accelerated by oncology clinical trials where examples of their use is more prevalent (12) however examples are emerging across other disease areas (13).

The International Council for Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) E9 guidelines state that 'although the early phases of drug development consist mainly of clinical trials that are exploratory in nature, statistical principles are also relevant'(14). As early phase clinical trials utilise statistical model based designs, the requirement for good quality SAPs, including additional statistical parameters and progression criteria to later phase research, becomes an even greater necessity (15,16), with the trial statistician playing a key role in designing, and undertaking analysis of early phase clinical trials.

Guidelines for the content of SAPs were published in 2017 (1) and highlighted the need for a detailed SAP to improve transparency, clinical trial quality, and accuracy. These guidelines were developed with the primary intention of being applicable to the analyses of later-phase randomised controlled trials (RCTs) and acknowledged that despite some recommendations being transferable, specific consideration and guidance are needed for early phase clinical trials. These guidelines were discussed at a UK Clinical Research Collaboration (UKCRC) Registered Clinical Trials Unit (CTU) Network Statisticians' Operational Group meeting in April 2018, confirming that specific consideration and guidance for early phase clinical trials was an area of unmet need. This was based on the fact that early phase clinical trials are often not randomised, often use adaptive designs, and often otherwise have statistical considerations and requirements that are different in character from those of later-phase, RCTs. This discussion led to this extension of those 2017 guidelines to address the needs and considerations of SAPs for early phase clinical trials. Given the drug development

pathway, early phase clinical trials are more prevalent than late phase (17), highlighting the importance and impact of this guidance.

Here we describe the development of an extension to published guidelines for SAP content to broaden their applicability to early phase clinical trials. These recommendations are intended to guide the authoring of SAPs for all early phase studies, irrespective of the study design used (rule-based, model-based, model-assisted, or randomised phase I trials; or single arm phase II designs). Beyond the scope of this extension are randomised phase II trials given that they are covered by the existing SAP guidelines for randomised trials.

Development of Early Phase SAP Guidance Extension

This guidance document encapsulates the findings of: a comprehensive search to identify existing published protocols, SAPs and SAP guidance; a survey of clinical trial funders and regulators; a survey of current practice of statisticians within UKCRC registered CTUs; a critical appraisal and expert review meeting; and a pilot of the proposed guidelines.

An overview of this process is given in appendix 1.

Literature Review of Existing Guidance

A literature review was undertaken to identify peer-reviewed publications of applicable guidelines, and example clinical trial protocols and SAPs, the search terms are given in appendix 2. The Enhancing the QUAlity and Transparency Of health Research (EQUATOR) network repository was searched to identify existing guidance; PubMED was searched to identify published SAPs; and a PubMED search of early phase (encompassing phase I and phase II) clinical trial protocols. The search of protocols was undertaken to capture the statistical detail contained within these documents, as SAPs may not always have been written for some early phase clinical trials.

These searches were performed independently by two statisticians, with all papers categorised as either a relevant and methodological paper; a relevant and published protocol or SAP, or; not relevant. Reasons for classifying publications as not relevant included the paper pertaining to: late phase trials; cluster-randomised trials; epidemiological studies; meta-analyses and systematic reviews; results papers; editorial publications, or; other reasons of non-relevance decided at the discretion of the reviewers (e.g. methodological papers, and SAPs for literature reviews). Any discrepancies regarding relevance categorisation were discussed and resolved by mutual agreement.

The literature review was initially performed in November 2019 and updated in October 2020. Of the 610 papers returned by the literature review, 500 were excluded due to non-relevance. The number of papers categorised according to each exclusion reason can be found in Figure 1. The exclusion reasons categorised as other included: statistical and clinical trials methodology and reviews; SAPs for sub-studies of RCTs; and SAPs for sub-studies. The SAP and protocol search identified two papers perceived to be SAP guidelines, the first being the original guidelines for the content of SAPs (1), and the second an overview to the typical content of SAPs for various study types (e.g., RCTs and observational studies) rather than proposing recommended content for SAPs (19). The literature review found three early phase SAP publications (20–22) and 105 published early phase protocols, containing some statistical content. The three published early phase SAPs were all single arm phase II clinical trials, and of the 105 published protocols one included a SAP as an appendix (23), seven indicated a separate SAP had been written (but did not make it available) and the remainder typically contained varying but limited statistical content. These findings are

consistent with the notion that while there has been an increase in the publication of SAPs, these are overwhelmingly late phase clinical trials.

Survey of clinical trial funders and regulators

The same clinical trial funders and regulators contacted during the original SAP guidance development (1) were contacted via email in January 2020. Funders were initially contacted to gauge whether they fund early phase clinical trials. If a response was not received, up to two further reminder emails were sent. Surveys were sent to all regulators and those organisations who confirmed the scope of their funding considerations would extend to early phase clinical trials. Consultation with clinical trial funders led to the identification of two additional dedicated early phase clinical trial funders who were also approached. A list of organisations contacted is provided in Appendix 3.

The goal of these surveys was to ascertain funding and regulatory requirements of design, analysis, publication, and SAP contents for early phase clinical trials.

Of the 39 funding institutions contacted, 28 responded to our request for information, 15 of whom indicated that they would fund early phase research. The European Medicines Agency (EMA) and Medicines and Healthcare products Regulatory Agency (MHRA) provided their regulatory requirements for early phase clinical trials. Additionally, guidance pertaining to the running and conduct of early phase clinical trials (including: multiple ICH documents (14,24–26); the Royal Statistical Society (RSS) working party report on statistical issues in First-in-Man studies (27); various Consolidated Standards of Reporting Trials (CONSORT) statements (28–31); and publicly available regulatory guidance (32)) were reviewed. The predominant documents funders referred to were ICH E9 (14) and existing late phase guidance (1). The prevailing findings were that:

- Dose-escalation decisions, stopping criteria, and interim go/no-go criteria are often poorly documented potentially resulting in ambiguous and non-robust decisions to escalate;
- The statistical design and analysis plan are often not clearly justified;
- Where model-based approaches are used, the choice of model should be clearly justified, and the risk of overdosing must be quantified and justified to be acceptable, supported by simulation where applicable;
- Appropriate sample sizes for early phase clinical trials can be better justified statistically, e.g. by simulation, as opposed to by mere custom and historical practice, and that sufficient detail regarding the sample size should be included (with supplemented code to help facilitate this where appropriate) to allow for full replication;
- The programming codes for modelling should either be suitably referenced or made available in the SAP so that escalation decisions can be replicated and reproduced, and;
- Standard statistical principles (e.g., the implication of interim analyses on the overall integrity of the clinical trial and Type I error control) are still applicable.

Survey of Clinical Trials Units

The survey was developed based on the original SAP guidance survey (1) and tailored to early phase clinical trials. The aim was to identify CTUs conducting early phase clinical trials and the current practice within those units for developing SAPs. The survey (Appendix 3) was circulated to CTUs in the UKCRC network. A list of the 53 registered CTUs was accessed (May 2020) from the UKCRC website (15) to cross-check responders.

A senior statistician at each CTU was asked to complete the survey to reflect practices and majority opinion within the statistician's CTU in May 2020. If no response was received then two reminder emails were sent via the mailing list to encourage responses, and then contacts at the unit were approached directly for a response.

Example SAPs shared by CTUs conducting early phase clinical trials were collated and reviewed for content to establish the current level of detail provided. To ensure as much coverage for study design types and disease areas, examples were sought from multiple scenarios, including design based (e.g. rule-based, model-based, or single-arm phase II) and disease based (oncology or non-oncology).

Of the 53 CTUs, 40 (75%) responded to the survey about their experiences of SAPs for early phase clinical trials. Of the 40 responders, 21 (53%) declared to design and analyse early phase clinical trials. On enquiry, the remaining 13 CTUs who did not respond did not conduct early phase clinical trials and so were not pursued further.

The prevailing practice of the 21 CTUs statistically designing and conducting early phase clinical trials at the time of survey circulation was to have a generic template (or set of instructions) applicable to all phases of clinical trial, without specific instructions or sections for early phase clinical trials (n=15, 73%). Three CTUs (14%) reported having no template. Importantly, only 3 (14%) CTUs reported having a template specific to early phase clinical trials. Most units (n=20, 95%) highlighted the desire for early phase SAP guidance, and all indicated they would use the guidance extension if it existed.

Critical Appraisal and Expert Review Meeting

The first draft of the guidance was produced using the literature reviews, funding and regulatory requirements, CTU examples, and authors' experience. An international expert review panel was convened of UK and US academic, pharmaceutical, NHS (UK only) and regulatory representatives. Details of the panel are provided in Appendix 4.

The expert review panel met virtually on the 26th October 2020 with contribution and attendance from 16 statisticians from 14 organisations. Participants critically appraised the first draft of the extension in advance of the meeting in preparation of wider discussion with the group. Considerable discussion centred on the level of detail required from the guidelines and areas that required expansion or clarification. Comments were received from meeting attendees, and a further five statisticians who were unable to attend the meeting.

Consensus was reached at this meeting regarding several areas for inclusion of content and recommendations. Following incorporation of comments, attendees reviewed and provided feedback prior to finalising the draft guidance extension for piloting within CTUs.

Piloting of the Early Phase Trial Extension

Following the expert review meeting, the guidelines extension was updated and piloted at six UK CTUs. The aim of the pilot was to ensure the guidance extension produced was fit-for-purpose, appropriate to the needs of statisticians authoring early phase clinical trial SAPs, and to identify any items requiring further clarification. The pilot was conducted between December 2020 to March 2021 whereby participating CTUs were invited to give formal feedback via email.

Piloting covered both phase I and phase II clinical trial designs, and different therapeutic areas. Feedback was universally positive with minimal amendments (improved wording and clarifications) made to the guidelines extension, which were then finalised.

Ethics

Consistent with the development of the original SAP guidance, ethical approval was not sought for the distribution of surveys, rather consent to take part was indicated by survey participation.

Early Phase Extension to Existing SAP Guidance

The resulting recommendations for the early phase clinical trial SAP content guidelines extension are presented in Table 1. Of the original 55 items proposed in the original SAP content guidance (1), 30 items have remained unchanged, 25 have been modified to better reflect early phase clinical trials, and a further 11 new items have been proposed. Significant modifications and new items include:

- Increased details regarding statistical design methodology, and where appropriate model choice;
- Update of outcome definitions to include definition of estimands in line with the principles outlined in ICH E9 (R1);
- Inclusion of simulation reports incorporating operating characteristics, to justify statistical design and/or sample size where applicable;
- Inclusion of code required for novel methodology;
- Inclusion of dose transition pathways, where appropriate, and;
- Amendments to wording to be more neutral to both frequentist and Bayesian methodology, to reflect that some early phase clinical trials designs, particularly phase I, are underpinned by Bayesian methods.

Minor changes were made to items including updates to the descriptions to ensure pertinence to early phase clinical trials. Items where there is no text in the “Recommended Early Phase Clinical Trials Extension” column in Table 1 indicates that the original item is appropriate and also covers the necessary content for early phase clinical trials.

An elaboration of each item within the extension guidelines is included in Appendix 5, with example text covering various early phase clinical trial designs and therapeutic areas. These examples are intended to be illustrative and are not an endorsement of the methods described.

Discussion

Critical appraisal of clinical trials is only possible if their design, conduct, and analysis are pre-defined, thoroughly and clearly described. It is crucial that planned trial analyses are suitably pre-defined, typically in a SAP. This increases the credibility of results by minimising the opportunity of making data-driven decisions, or selecting estimands or methodology to produce a more positive trial outcome. With increased focus on transparency and concerns regarding reproducibility of results alongside the ability to reconstruct clinical trial design and analyses, guidance has been produced for prospective reporting of SAPs for RCTs (33) which advocate that SAPs should be made publicly available. However, this has generally focused on late phase clinical trials and the needs of early phase clinical trials have not explicitly been considered, acknowledging the need for extensions (1). The absence of transparency for early phase clinical trials and guidance tailored to these designs could result in biased results, which could in turn misinform decision-making in clinical development. This SAP guidance extension has been developed to enable statisticians, triallists, and clinical investigators conducting early phase clinical trials to author clear and relevant SAPs for those clinical trials. The desired outcome being that these clinical trials would be run with increased transparency and increase the likelihood of accurate conclusions.

One of the major updates of the extension is the development of wording regarding outcome measures and estimands (Appendix 5, Section 6, item 26), following ICH E9 (R1) addendum on estimands and sensitivity analyses in clinical trials (26). Although ICH E9 (R1) focuses on the analysis and interpretability of late-phase trials, it clearly states that the same principles apply to single arm trials, therefore encompassing early phase designs. At the time of publication, estimands are not widely used in early phase clinical trials with working groups convened specifically to advise on this issue (for example, the Oncology Estimand Working Group (34) and the Estimand Implementation Working Group (35)). These groups will provide or publish guidance and examples pertaining to estimands for early phase clinical trials, which should be incorporated into the authoring of SAPs in the future. This may also necessitate revisions to this document.

Another significant component to the extension includes a simulation report template. In all phases of clinical trials there is a need to ensure that the clinical trial will yield an accurate, unbiased result. For fixed or late phase clinical trials, this typically manifests as a formal sample size calculation. However, for model-assisted and model-based early phase designs, this is not appropriate. Instead, simulations to assess the designs operating characteristics are needed to ensure the clinical trial will yield a result and provide sufficient overdose control (36). A template for the suitable simulation report has not been developed as part of this project as they depend on multiple variables including disease area, trial question, trial design, and methodology. Instead, suggestions for content are provided in Appendix 5, Item 33. The same is true for relevant code and a reports template, with suggestions provided in Appendix 5, items 35 and 36 respectively.

In early phase clinical trials, critical decisions regarding the trial progression often need to be made at multiple time points, potentially as early as after the first patient has completed a specified evaluation period. Therefore, compared to late phase clinical trials, SAPs for early phase clinical trials will generally need to be authored earlier. It was acknowledged at the expert review meeting that the first version of the SAP should be signed off prior to the trial opening, but acknowledging this is not always feasible, and the panel recommended finalising the first version of the SAP prior to the first analysis of clinical trial data, for example prior to evaluation of a potential first dose escalation or any interim analysis.

It was acknowledged that the SAP should encompass all relevant points from Table 1, but recommended that signposting be used to indicate location of details that are captured elsewhere in documents such as in the clinical trial protocol or a simulation report, to avoid replication. The SAP is not a standalone document and should be read in conjunction with other trial-related documentation.

In certain scenarios, for example rule-based designs, it has been proposed that a SAP may not be required for early phase clinical trials, for instance if all the applicable content according to these guidelines is sufficiently detailed in the protocol. However, it is worth noting that any minor change to that content as the clinical trial progresses would then necessitate a protocol amendment. The combination of running an efficient clinical trial and the level of detail appropriate for analysis of early phase clinical trials provided within this content justifies a standalone SAP.

Whilst not all UKCRC registered CTUs responded to our survey, this extension captures the opinions of all those who design and analyse early phase clinical trials. Our work built upon the original SAP guidelines which included a Delphi survey (37). In developing the extension to early phase clinical trials an additional Delphi Survey was not considered necessary as the aim was to build upon existing knowledge rather than requiring repetition of the process. Instead, we ensured that all relevant stakeholder groups, including CTUs, funders and regulators, were included in the elicitation of

information with the expert multidisciplinary panel assuming responsibility of ensuring detailed considerations of the produced guidelines.

Our expert panel comprised of more academic statisticians compared to regulatory or pharmaceutical statisticians. This under-representation was identified and more representatives from these areas were invited to participate on the panel. In addition, some provided comments and responded to surveys but were not able to participate in the expert panel. Importantly, every expert panel member was invited to provide their opinions outside of the meeting and review the proposed guideline extension independently and share with their colleagues. When these views were combined, all views were weighted equally regardless of further participation in the extension's development, and consideration was taken in ensuring that the opinions of all stakeholder groups were reflected in the final version, which all authors have contributed to and agreed upon.

Given the multitude and complexity of designs for early phase clinical trials, this guidance was developed to be as generic and applicable as possible across all designs. A proportionate approach was taken, striking a balance between an increased number of items but not increasing beyond the minimum number of items needed to cover adequately the diverse features of early phase clinical trial designs and analysis requirements. Methodological developments within early phase clinical trials are increasingly prevalent and as such continued progress in this area is expected. The guidance may need to be updated when developments in statistical techniques emerge, and as such it may be useful to periodically review methodology and guidance documentation around early phase clinical trials for necessary updates.

There remain certain types of trials, such as Bayesian or adaptive trials, not covered by this extension or the original guidance, which will require additional considerations incorporating available regulatory and published guidance.

This guidance document provides a necessary extension to the SAP guidelines paper published by Gamble *et al*. (1) for early phase clinical trials.

Conclusions

In conclusion, guidelines are presented here for an extension to existing SAP content guidance appropriate for early phase clinical trials. Adherence to this extended guidance will support those working in early phase clinical trials in producing robust conclusions to ensure correct decisions are taken forward.

Acknowledgments

This project is funded by the National Institute for Health Research (NIHR) CTU Support Funding scheme. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

This project was supported and endorsed by the UK Clinical Research Collaboration Registered Clinical Trials Unit Network.

The authors would also like to thank: Peter Brocklehurst and Birmingham Clinical Trials Unit (BCTU) for their assistance in the initiation and support of this project; all UKCRC registered CTUs who responded to our surveys and provided example documents; the various clinical trials units (The Institute of Cancer Research (ICR), Cambridge Clinical Trials Unit (CCTU), Leeds Clinical Trials Research Unit (CTRU), Royal Marsden Clinical Trials Unit (RM-CTU), the Centre for Statistics in Medicine at the University of Oxford (CSM), and Cancer Research Clinical Trials Unit (CRCTU)) and

the individuals at these units (Muneeb Mahmud and Jan Rekowski (ICR), Simran Vaja (CCTU), Amy Coulson (CTRU), Catey Bunce and Kabir Mohammed (RM-CTU), Cathy Qi (CSM), and Amit Patel and Shanna Maycock (CRCTU)) for piloting the guidelines; Louise Stanton and Louise Williams for facilitating communication with the UKCRC CTU Network; Siân Lax (CRCTU) for her help in preparing this manuscript; and all individuals from funding and regulatory institutions who have responded to our surveys.

Staff at the CRCTU are supported by a core funding grant from Cancer Research UK (C22436/A25354 and C22436/A28028), and further grants from Birmingham Biomedical Research Centre (BRC; funded by the National Institute for Health Research (NIHR)) and Advanced Therapy Treatment Centre (ATTC; funded by Innovate UK).

Word count (4086)

Contributors and sources:

PG, CY, SB, DS, and CG authored and were awarded the grant funding. VH and PG performed the literature review, and authored the: survey, guidance, and manuscript. All authors attended the expert review meeting or provided comments, contributed to the underpinning research, and subsequent drafts of the guidance and manuscript, and approved the final version. VH and PG act as guarantors. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.There is no additional data available.

Provenance and peer review:

Not commissioned; externally peer reviewed

References

1. Gamble C, Krishan A, Stocken D, et al. Guidelines for the content of statistical analysis plans in clinical trials. JAMA. 2017 Dec 19;318(23):2337-2343. doi: 10.1001/jama.2017.18556. PMID: 29260229.
2. Storer BE. Design and Analysis of Phase I Clinical Trials. Biometrics. 1989;45(3):925–37. PMID: 2790129.
3. Oron AP, Azriel D, Hoff PD. Dose-finding designs: The role of convergence properties. Int J Biostat. 2011 Oct 27;7:Article 39. doi: 10.2202/1557-4679.1298. PMID: 22718676
4. Love SB, Brown S, Weir CJ, et al. Embracing model-based designs for dose-finding trials. Br J Cancer. 117, 332–339 (2017). https://doi.org/10.1038/bjc.2017.186
5. O'Quigley J, Pepe M, Fisher L. Continual Reassessment Method: A Practical Design for Phase 1 Clinical Trials in Cancer. Biometrics. 1990 Mar;46(1):33-48. PMID: 2350571.
6. Babb J, Rogatko A, Zacks S. Cancer phase I clinical trials: Efficient dose escalation with overdose control. Stat Med. 1998 May 30;17(10):1103-20. doi: 10.1002/(sici)1097-0258(19980530)17:10<1103::aid-sim793>3.0.co;2-9. PMID: 9618772.
7. Ying Kuen Cheung, Chappell R. Sequential designs for phase I clinical trials with late-onset toxicities. Biometrics. 2000 Dec;56(4):1177-82. doi: 10.1111/j.0006-341x.2000.01177.x.

PMID: 11129476.

8. Neuenschwander B, Branson M, GsPoner T. Critical aspects of the Bayesian approach to phase I cancer trials. Stat Med. 2008 Jun 15;27(13):2420-39. doi: 10.1002/sim.3230. PMID: 18344187.

9. Ji Y, Wang SJ. Modified toxicity probability interval design: A safer and more reliable method than the 3 + 3 design for practical phase I trials. J Clin Oncol. 2013 May 10;31(14):1785-91. doi: 10.1200/JCO.2012.45.7903. Epub 2013 Apr 8. PMID: 23569307; PMCID: PMC3641699.

10. Iasonos A, O'Quigley J. Randomised Phase 1 clinical trials in oncology. Br J Cancer. 2021. https://doi.org/10.1038/s41416-021-01412-y

11. Brown SR, Gregory WM, Twelves C, Brown J. A Practical Guide to Designing Phase II Trials in Oncology. A Practical Guide to Designing Phase II Trials in Oncology. 2014. ISBN: 978-1-118-57090-6

12. Iasonos A, Wilton AS, Riedel ER, Seshan VE, Spriggs DR. A comprehensive comparison of the continual reassessment method to the standard 3 + 3 dose escalation scheme in Phase I dose-finding studies. Clin Trials. 2008;5(5):465-77. doi: 10.1177/1740774508096474. PMID: 18827039; PMCID: PMC2637378.

13. Cole M, Yap C, Buckley C, et al. TRAFIC: statistical design and analysis plan for a pragmatic early phase 1/2 Bayesian adaptive dose escalation trial in rheumatoid arthritis. Trials 22, 433 (2021). https://doi.org/10.1186/s13063-021-05384-5

14. European Medicines Agency. ICH E9 - Statistical Principles for Clinical Trials. ICH Guidel. 2006;

15. Yap C, Billingham LJ, Cheung YK, Craddock C, O'Quigley J. Dose transition pathways: The missing link between complex dose-finding designs and simple decision-making. Clin Cancer Res. 2017 Dec 15;23(24):7440-7447. doi: 10.1158/1078-0432.CCR-17-0582. Epub 2017 Jul 21. PMID: 28733440.

16. Wheeler GM, Mander AP, Bedding A, et al. How to design a dose-finding study using the continual reassessment method. BMC Med Res Methodol. 2019 Jan 18;19(1):18. doi: 10.1186/s12874-018-0638-z. PMID: 30658575; PMCID: PMC6339349.

17. Gresham G, Meinert JL, Gresham AG, Meinert CL. Assessment of Trends in the Design, Accrual, and Completion of Trials Registered in ClinicalTrials.gov by Sponsor Type, 2000-2019. JAMA Netw Open. 2020;

18. UK Clinical Research Collaborative Registered Clinical Trials Units. 2019/20 UKCRC Registration ID Numbers [Internet]. Available from: https://ukcrc-ctu.org.uk/wp-content/uploads/2021/04/2019-20_reg_ids_nov19.pdf

19. Yuan I, Topjian AA, Kurth CD, et al. Guide to the statistical analysis plan. Paediatr Anaesth. 2019; 29:237–242. https://doi.org/10.1111/pan.13576

20. Blair GW, Appleton JP, Law ZK, et al. Preventing cognitive decline and dementia from cerebral small vessel disease: The LACI-1 Trial. Protocol and statistical analysis plan of a phase IIa dose escalation trial testing tolerability, safety and effect on intermediary endpoints of isosorbide mon. Int J Stroke. 2018 Jul;13(5):530-538. doi: 10.1177/1747493017731947. Epub 2017 Sep 14. PMID: 28906205

21. Chiodini P, Arenare L, Piccirillo MC, Perrone F, Gallo C. A phase 2, open label, multicenter, single arm study of tocilizumab on the efficacy and tolerability of tocilizumab in the treatment of patients with COVID-19 pneumonia (TOCIVID-19 trial): Statistical analysis plan.

Contemp Clin Trials Commun. 2020 Dec;20:100665. doi: 10.1016/j.conctc.2020.100665. Epub 2020 Oct 7. PMID: 33043164; PMCID: PMC7538865.

22. van der Westhuizen A, Knoblauch N, Graves MC, Levy R, Vilain RE, Bowden NA. Pilot early phase II study of decitabine and carboplatin in patients with advanced melanoma. Medicine (Baltimore). 2020; 99(25):e20705. doi:10.1097/MD.0000000000020705.

23. Ramanan A V., Dick AD, Jones AP, et al. A phase II trial protocol of Tocilizumab in anti-TNF refractory patients with JIA-associated uveitis (the APTITUDE trial). BMC Rheumatol 2, 4 (2018). https://doi.org/10.1186/s41927-018-0010-2

24. Eurpoean Medicines Agency. ICH E3 - Structure and Content of Clinical Study Reports. ICH Guidel. 1996;

25. European Medicines Agency. ICH E4 - Dose Response Information to Support Drug Registratione. ICH Guidel. 1994;

26. European Medicines Agency. ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. ICH Guidel. 2020;

27. Royal Statistical Society. Report of the Working Party on Statistical Issues in First-in-Man studies. 2007.

28. Schulz KF, Altman DC, Moher D. CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. Ital J Public Health. 2010 Mar 23;340:c332. doi: 10.1136/bmj.c332. PMID: 20332509; PMCID: PMC2844940.

29. Ioannidis JPA, Evans SJW, Gøtzsche PC, et al. Better reporting of harms in randomized trials: An extension of the CONSORT statement. Ann Intern Med. 2004 Nov 16;141(10):781-8. doi: 10.7326/0003-4819-141-10-200411160-00009. PMID: 15545678.

30. Vohra S, Shamseer L, Sampson M, et al. CONSORT extension for reporting N-of-1 trials (CENT) 2015 Statement. J Clin Epidemiol. 2015 May 14;350:h1738. doi: 10.1136/bmj.h1738. Erratum in: BMJ. 2016 Oct 10;355:i5381. PMID: 25976398.

31. Eldridge SM, Chan CL, Campbell MJ, et al. CONSORT 2010 statement: Extension to randomised pilot and feasibility trials. Pilot Feasibility Stud. 2016 Oct 21;2:64. doi: 10.1186/s40814-016-0105-8. PMID: 27965879; PMCID: PMC5154046.

32. Rantell K, on behalf of the Medicines and Healthcare products Regulatory Agencey. Regulatory perspective of early phase I/II designs, including basket, umbrella and platform designs. 2019;

33. Hemming K, Kearney A, Gamble C, et al. Prospective reporting of statistical analysis plans for randomised controlled trials. Trials 21, 898 (2020). https://doi.org/10.1186/s13063-020-04828-8

34. Oncology Estimand Working Group. Oncology Estimand Working Group: A cross-industry international working group [Internet]. 2021 [cited 2021 Jul 15]. Available from: http://www.oncoestimand.org/

35. EFSPI/EFPIA Estimands Implementation Working Group. Estimands Implementation Working Group [Internet]. 2021 [cited 2021 Jul 22]. Available from: https://www.efspi.org/EFSPI/Working_Groups/EFSPI_EFPIA_EIWG.aspx

36. Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. Stat Med. 2019 May 20;38(11):2074-2102. doi: 10.1002/sim.8086. Epub 2019 Jan 16. PMID: 30652356; PMCID: PMC6492164.

37. Rowe G, Wright G. The Delphi technique as a forecasting tool: Issues and analysis. Int J Forecast. 1999; 15. Pages 353-375, ISSN 0169-2070, https://doi.org/10.1016/S0169-2070(99)00018-7

Table 1: SAP Content Guidance Extension for Early Phase Clinical Trials

	Original SAP Content Guidance		Recommended Early Phase Clinical Trials Extension Guidance	
Section/Item	**Item No**	**Description**	**Item No**	**Description**
Section 1: Administrative Information				
Title and trial registration	1a	Descriptive title that matches the protocol, with SAP either as a forerunner or subtitle, and trial acronym (if applicable)		
	1b	Trial registration number		
SAP version	2	SAP version number with dates		
Protocol version	3	Reference to version of protocol being used		
SAP revisions	4a	SAP revision history		
	4b	Justification for each SAP revision		
	4c	Timing of SAP revisions in relation to interim analyses, etc.		
Roles and responsibility	5	Names, affiliations, and roles of SAP contributors		
Signatures of:	6a	Person writing the SAP		
	6b	Senior statistician responsible		
	6c	Chief investigator/clinical lead		
Section 2: Introduction				
Background and rationale	7	Synopsis of trial background and rationale including a brief description of research question and brief justification for undertaking the trial		
Objectives	8	Description of specific objectives or hypotheses	8	Description of specific question, objectives or hypotheses. It should be made clear what the key objectives are (for example primary and secondary objectives that encompasses toxicity, efficacy, PK, PD, or some combination).
Section 3: Study Methods				
Trial design	9	Brief description of trial design including type of trial (e.g., parallel group, multiarm, crossover, factorial) and allocation ratio and may include brief description of interventions	9a	Brief description of trial design, including the trial phase and the design method (dose escalation e.g., CRM or single-arm phase II e.g., Simon's Two Stage). If the trial has a randomised element to it, summary information regarding the randomisation, including the allocation ratio, should be specified.
			9b	Treatment information, including the dose levels of intervention(s). Where appropriate, and if multiple doses are used, the following should also be reported: the ordering and combination (in the instance of multiple agents under investigation) of dose levels, and the dose level to start at.

			9c	Details regarding the statistical methodology underpinning the trial, including the choice of the number of parameters in the model if applicable, its empirical form and all formulae. It is also important to ensure all model parameters are given, including where appropriate, the weights of the model.
			9d	Rules of the trial design and model. Here information on the target objective (toxicity, response, PK, or PD, either singularly or in combination), classification of overdosing, and any stopping boundaries should be given. This may include the desired certainty in these estimates. Moreover, where dose decisions (e.g. escalation, de-escalation, remain at current dose or stop early) are to occur, details regarding dose transitions and dose skipping should be given.
			9e	Experimental details and design specifics. For dose escalation trials, information regarding cohort size, including whether this is fixed or flexible should be given. Indication of the stopping rules for interim and final evaluations. For model-based and model-assisted designs, details on the prior including full skeleton (if applicable) and its elicitation should be given. For single arm phase II trials, the target sample size and, where appropriate, the timing of any interim analyses
Randomisation	10	Randomisation details, e.g., whether any minimization or stratification occurred (including stratifying factors used or the location of that information if it is not held within the SAP)	10	Where appropriate, randomisation details e.g., whether any minimisation or stratification occurred (including stratifying factors used or the location of that information if it is not held within the SAP) and where applicable, details on blinding.
Sample size	11	Full sample size calculation or reference to sample size calculation in protocol (instead of replication in SAP)	11	Full sample size determination or justification or reference to relevant section in protocol (instead of replication in SAP)
Framework	12	Superiority, equivalence, or noninferiority hypothesis testing framework, including which comparisons will be presented on this basis	12	If applicable, specify whether trial is to be performed under hypothesis testing or Bayesian framework.
Statistical interim analyses and stopping guidance	13a	Information on interim analyses specifying what interim analyses will be carried out and listing of time points	13a	Information pertaining to interim dose decisions (e.g. escalation, de-escalation, remain at current dose or stop early).
	13b	Any planned adjustment of the significance level due to interim analysis	13b	Information on other interim analyses specifying what and when interim analyses will be conducted.
	13c	Details of guidelines for stopping the trial early	13c	Any planned adjustment of the significance level due to interim analysis
			13d	Details of guidelines for stopping the trial early
Timing of final analysis	14	Timing of final analysis, e.g., all outcomes analysed collectively or timing stratified by planned length of follow-up		
Timing of outcome assessments	15	Time points at which the outcomes are measured including visit "windows"		
Section 4: Statistical Principles				

Indications of uncertainty [a]	16	Level of statistical significance	16*	Level of statistical significance
	17	Description and rationale for any adjustment for multiplicity and, if so, detailing how the type 1 error is to be controlled	17	Description of any planned adjustment for multiplicity, and if so, including how the type 1 error is to be controlled
	18	Confidence intervals to be reported	18	Either confidence or credible intervals to be reported (appropriately picked dependent on the trial methodology).
Adherence and protocol deviations	19a	Definition of adherence to the intervention and how this is assessed including extent of exposure		
	19b	Description of how adherence to the intervention will be presented		
	19c	Definition of protocol deviations for the trial		
	19d	Description of which protocol deviations will be summarized		
Analysis populations	20	Definition of analysis populations, e.g., intention to treat, per protocol, complete case, safety.	20	Clear definition of the trial/dose cohort(s) including how cohorts will be referred to, how patients enter cohorts, the minimum number of patients needed to be in a cohort (and how long they have been in) before dose escalation decisions can be made. Trial level definitions of patient populations (e.g., per-protocol, intention to treat, safety) should also be given. Details regarding evaluable patients and specify what happens to unevaluable patients should also be made. These definitions should also be provided for any interim analysis populations.
Section 5: Trial Populations				
Screening data	21	Reporting of screening data (if collected) to describe representativeness of trial sample		
Eligibility	22	Summary of eligibility criteria		
Recruitment	23	Information to be included in the CONSORT flow diagram		
Withdrawal/follow-up	24a	Level of withdrawal, e.g., from intervention and/or from follow-up		
	24b	Timing of withdrawal/lost to follow-up data		
	24c	Reasons and details of how withdrawal/lost to follow-up data will be presented		
Baseline patient characteristics	25a	List of baseline characteristics to be summarized	25a*	List of baseline characteristics to be summarised
	25b	Details of how baseline characteristics will be descriptively summarized		
Section 6: Analysis				
Estimand definition [b]		List and describe each primary and secondary outcome including details of:		List and describe each primary and secondary estimands including details of:
	26a	Specification of outcomes and timings. If applicable include the order of importance of primary or key secondary end points (e.g., order in which they will be tested)	26a	Details of the treatment (including treatment combinations), and any alternative treatments to which comparisons will be made (where appropriate). For dose-finding trials, information on whether analysis will be performed per cohort, per dose received, pooled across all dose levels, or in some combination of these

	26b	specific measurement and units (e.g., glucose control, hbA1c [mmol/mol or %])	26b	The trial population, defined with reference to item 20, pertinent to each estimand
	26c	Any calculation or transformation used to derive the outcome (e.g., change from baseline, QoL score, time to event, logarithm, etc.)	26c	The variable of interest to be obtained for each patient that is required to address the scientific question. For outcomes recorded at multiple time points, distinction as to which of these time points are required for the estimand
			26d	Intercurrent events and their handling strategy, including adjustment to analysis
			26e	Detail the population–level summary measure for each estimand
Analysis methods	27a	what analysis method will be used and how the treatment effects will be presented	27a	What estimator and analysis method will be used and how the results will be presented
	27b	any adjustment for covariates	27b*	Any adjustments for covariates
	27c	methods used for assumptions to be checked for statistical methods	27c*	Methods used to check assumptions of the underlying statistical methods and goodness of fit for the model
	27d	details of alternative methods to be used if distributional assumptions do not hold, e.g., normality, proportional hazards, etc.	27d*	Details of alternative methods to be used if distributional assumptions do not hold
	27e	any planned sensitivity analyses for each outcome where applicable	27e	Any planned sensitivity analyses for each estimand where applicable
	27f	any planned subgroup analyses for each outcome including how subgroups are defined	27f	Any planned subgroup analyses for each estimand including how subgroups are defined
Missing data	28	Reporting and assumptions/statistical methods to handle missing data (e.g., multiple imputation)	28*	Reporting and assumptions/statistical methods to handle missing data (e.g., multiple imputation)
Additional analyses	29	Details of any additional statistical analyses required, e.g., complier-average causal effect analysis		
Harms	30	Sufficient detail on summarizing safety data, e.g., information on severity, expectedness, and causality; details of how adverse events are coded or categorized; how adverse event data will be analysed, i.e., grade 3/4 only, incidence case analysis, intervention emergent analysis	30	Sufficient detail on summarizing safety data outside of that used for dose escalation (e.g., non-DLT safety data), e.g., information on severity, expectedness, and causality; details of how adverse events are coded or categorised; how adverse event data will be analysed, i.e., by grade, incidence case analysis, intervention emergent analysis
Statistical software	31	Details of statistical packages to be used to carry out analyses	31	Details of statistical packages to be used to carry out design, simulation and analyses
References	32a	References to be provided for nonstandard statistical methods		
	32b	Reference to Data Management Plan		
	32c	Reference to the Trial Master File and Statistical Master File		
	32d	Reference to other standard operating procedures or documents to be adhered to		
Section 7: Suggested SAP Appendices				
Simulation Report			33	Operating characteristics of the trial design to assess the probability of trial success under different plausible scenarios.
Dose transition Pathways			34	For dose-escalation trials, indication of the dose transition pathways (either using tables or trees/graphs) under different DLT scenarios.

Code			35	Full model specification and programming code used for evaluation of dose-escalation decisions
Reports Template			36	Optional section detailing exemplar tables, graphs and report templates.

Notes:

a. This item was called 'Confidence intervals and P values' in the Gamble *et al.* paper (1)]. It has been changed to 'Indications of uncertainty' to reflect that many early phase trials designs are underpinned by Bayesian methodology.

b. This item was called 'Outcome definitions' in the Gamble *et al.* paper (1). It has been changed to 'Estimand definitions' following the wider adoption of ICH E9 (R1; Addendum on estimands).

* These represent items for which the table description has remained the same, but the explanation, as detailed in Appendix 5, has been amended.

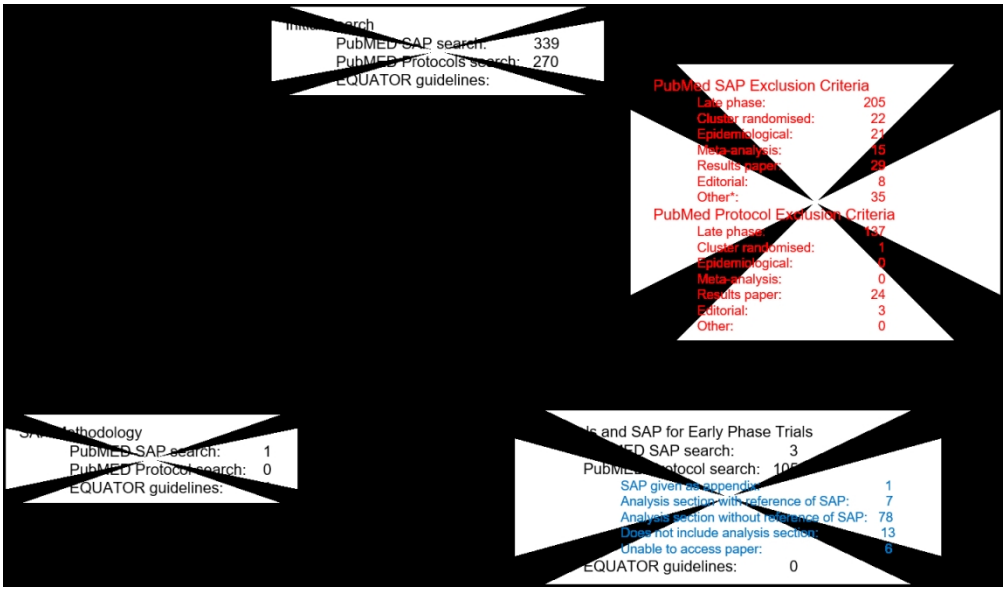
Items where there is no text in the table for the extensions for early phase column indicates that the original item is appropriate and covers the necessary content for early phase trials with examples for all items given in Appendix 5.

Abbreviations:

- CONSORT: CONsolidated Standards Of Reporting Trial
- CRM: Continual Reassessment Method
- DLT: Dose Limiting Toxicity
- PD: PharmacoDynamics
- PK: PharmacoKinetics
- SAP: Statistical Analysis Plan

Figure 1. Literature Review Results

Notes: *Correct as of 22nd October 2020*. *Other reasons of non-relevance include: Statistical and clinical trials methodology, and literature reviews; SAPs for sub-studies of randomised controlled trials; and SAPs for sub-studies.



Caption: Figure 1. Literature Review Results
Notes: Correct as of 22nd October 2020. *Other reasons of non-relevance include: Statistical and clinical trials methodology, and literature reviews; SAPs for sub-studies of randomised controlled trials; and SAPs for sub-studies.

613x358mm (59 x 59 DPI)

Checklist 1: SAP Content Guidance Extension for Early Phase Clinical Trials

	Recommended Early Phase Clinical Trials Extension Guidance		Page
Section/Item	**Item No**	**Description**	
Section 1: Administrative Information			
Title and trial registration	1a	Descriptive title that matches the protocol, with SAP either as a forerunner or subtitle, and trial acronym (if applicable)	
	1b	Trial registration number	
SAP version	2	SAP version number with dates	
Protocol version	3	Reference to version of protocol being used	
SAP revisions	4a	SAP revision history	
	4b	Justification for each SAP revision	
	4c	Timing of SAP revisions in relation to interim analyses, etc.	
Roles and responsibility	5	Names, affiliations, and roles of SAP contributors	
Signatures of:	6a	Person writing the SAP	
	6b	Senior statistician responsible	
	6c	Chief investigator/clinical lead	
Section 2: Introduction			
Background and rationale	7	Synopsis of trial background and rationale including a brief description of research question and brief justification for undertaking the trial	
Objectives	8	Description of specific question, objectives or hypotheses. It should be made clear what the key objectives are (for example primary and secondary objectives that encompasses toxicity, efficacy, PK, PD, or some combination).	
Section 3: Study Methods			
Trial design	9a	Brief description of trial design, including the trial phase and the design method (dose escalation e.g., CRM or single-arm phase II e.g., Simon's Two Stage). If the trial has a randomised element to it, summary information regarding the randomisation, including the allocation ratio, should be specified.	
	9b	Treatment information, including the dose levels of intervention(s). Where appropriate, and if multiple doses are used, the following should also be reported: the ordering and combination (in the instance of multiple agents under investigation) of dose levels, and the dose level to start at.	
	9c	Details regarding the statistical methodology underpinning the trial, including the choice of the number of parameters in the model if applicable, its empirical form and all formulae. It is also important to ensure all model parameters are given, including where appropriate, the weights of the model.	
	9d	Rules of the trial design and model. Here information on the target objective (toxicity, response, PK, or PD, either singularly or in combination), classification of overdosing, and any stopping boundaries should be given. This may include the desired certainty in these estimates. Moreover, where dose decisions (e.g. escalation, de-escalation, remain at current dose or stop early) are to occur, details regarding dose transitions and dose skipping should be given.	
	9e	Experimental details and design specifics. For dose escalation trials, information regarding cohort size, including whether this is fixed or flexible should be given. Indication of the stopping rules for interim and final evaluations. For model-based and model-assisted designs, details on the prior including full skeleton (if applicable) and its elicitation should be given. For single arm phase II trials, the target sample size and, where appropriate, the timing of any interim analyses	

Randomisation	10	Where appropriate, randomisation details e.g., whether any minimisation or stratification occurred (including stratifying factors used or the location of that information if it is not held within the SAP) and where applicable, details on blinding.	
Sample size	11	Full sample size determination or justification or reference to relevant section in protocol (instead of replication in SAP)	
Framework	12	If applicable, specify whether trial is to be performed under hypothesis testing or Bayesian framework.	
Statistical interim analyses and stopping guidance	13a	Information pertaining to interim dose decisions (e.g. escalation, de-escalation, remain at current dose or stop early).	
	13b	Information on other interim analyses specifying what and when interim analyses will be conducted.	
	13c	Any planned adjustment of the significance level due to interim analysis	
	13d	Details of guidelines for stopping the trial early	
Timing of final analysis	14	Timing of final analysis, e.g., all outcomes analysed collectively or timing stratified by planned length of follow-up	
Timing of outcome assessments	15	Time points at which the outcomes are measured including visit “windows”	
Section 4: Statistical Principles			
Indications of uncertainty [a]	16	Level of statistical significance	
	17	Description of any planned adjustment for multiplicity, and if so, including how the type 1 error is to be controlled	
	18	Either confidence or credible intervals to be reported (appropriately picked dependent on the trial methodology).	
Adherence and protocol deviations	19a	Definition of adherence to the intervention and how this is assessed including extent of exposure	
	19b	Description of how adherence to the intervention will be presented	
	19c	Definition of protocol deviations for the trial	
	19d	Description of which protocol deviations will be summarized	
Analysis populations	20	Clear definition of the trial/dose cohort(s) including how cohorts will be referred to, how patients enter cohorts, the minimum number of patients needed to be in a cohort (and how long they have been in) before dose escalation decisions can be made. Trial level definitions of patient populations (e.g., per-protocol, intention to treat, safety) should also be given. Details regarding evaluable patients and specify what happens to unevaluable patients should also be made. These definitions should also be provided for any interim analysis populations.	
Section 5: Trial Populations			
Screening data	21	Reporting of screening data (if collected) to describe representativeness of trial sample	
Eligibility	22	Summary of eligibility criteria	
Recruitment	23	Information to be included in the CONSORT flow diagram	
Withdrawal/follow-up	24a	Level of withdrawal, e.g., from intervention and/or from follow-up	
	24b	Timing of withdrawal/lost to follow-up data	
	24c	Reasons and details of how withdrawal/lost to follow-up data will be presented	
Baseline patient characteristics	25a	List of baseline characteristics to be summarised	
	25b	Details of how baseline characteristics will be descriptively summarized	
Section 6: Analysis			
Estimand definition [b]		List and describe each primary and secondary estimands including details of:	
	26a	Details of the treatment (including treatment combinations), and any alternative treatments to which comparisons will be made (where appropriate). For dose-finding trials, information on whether analysis will be performed per cohort, per dose received, pooled across all dose levels, or in some combination of these	
	26b	The trial population, defined with reference to item 20, pertinent to each estimand	

	26c	The variable of interest to be obtained for each patient that is required to address the scientific question. For outcomes recorded at multiple time points, distinction as to which of these time points are required for the estimand	
	26d	Intercurrent events and their handling strategy, including adjustment to analysis	
	26e	Detail the population–level summary measure for each estimand	
Analysis methods	27a	What estimator and analysis method will be used and how the results will be presented	
	27b	Any adjustments for covariates	
	27c	Methods used to check assumptions of the underlying statistical methods and goodness of fit for the model	
	27d	Details of alternative methods to be used if distributional assumptions do not hold	
	27e	Any planned sensitivity analyses for each estimand where applicable	
	27f	Any planned subgroup analyses for each estimand including how subgroups are defined	
Missing data	28	Reporting and assumptions/statistical methods to handle missing data (e.g., multiple imputation)	
Additional analyses	29	Details of any additional statistical analyses required, e.g., complier-average causal effect analysis	
Harms	30	Sufficient detail on summarizing safety data outside of that used for dose escalation (e.g., non-DLT safety data), e.g., information on severity, expectedness, and causality; details of how adverse events are coded or categorised; how adverse event data will be analysed, i.e., by grade, incidence case analysis, intervention emergent analysis	
Statistical software	31	Details of statistical packages to be used to carry out design, simulation and analyses	
References	32a	References to be provided for nonstandard statistical methods	
	32b	Reference to Data Management Plan	
	32c	Reference to the Trial Master File and Statistical Master File	
	32d	Reference to other standard operating procedures or documents to be adhered to	
Section 7: Suggested SAP Appendices			
Simulation Report	33	Operating characteristics of the trial design to assess the probability of trial success under different plausible scenarios.	
Dose transition Pathways	34	For dose-escalation trials, indication of the dose transition pathways (either using tables or trees/graphs) under different DLT scenarios.	
Code	35	Full model specification and programming code used for evaluation of dose-escalation decisions	
Reports Template	36	Optional section detailing exemplar tables, graphs and report templates.	

Notes:

a. This item was called 'Confidence intervals and P values' in the Gamble *et al.* paper (1)]. It has been changed to 'Indications of uncertainty' to reflect that many early phase trials designs are underpinned by Bayesian methodology.

b. This item was called 'Outcome definitions' in the Gamble *et al.* paper (1). It has been changed to 'Estimand definitions' following the wider adoption of ICH E9 (R1; Addendum on estimands).

Abbreviations:

- CONSORT: CONsolidated Standards Of Reporting Trial
- CRM: Continual Reassessment Method
- DLT: Dose Limiting Toxicity
- PD: PharmacoDynamics
- PK: PharmacoKinetics
- SAP: Statistical Analysis Plan

Appendix 1: Overview of Early Phase SAP Guidance Extension Process

Figure S1: Schema detailing overview of early phase SAP guidance extension process

Appendix 2: Literature Search Terms

PubMED SAP Search

1. "Statistical Analysis Plan"

PubMED Protocol search

1. "Phase I"
2. "Phase II"
3. 1 OR 2
4. "Protocol"
5. "Trial"
6. 4 AND 5
7. 3 AND 6

Appendix 3. Surveys of Clinical Trials Regulators, Funders and CTUs

List of Clinical Trial Regulators and Funders Contacted:

Name of Organisation	Type
Medicines & Healthcare Products Regulatory Agency (MHRA)	Regulator
European Medicines Agency (EMA)	Regulator
U.S. Food and Drug Administration (FDA)	Regulator
Health Technology Assessment (HTA)	Funder
Public Health Research (PHR)	Funder
Health Services and Delivery Research (HS&DR)	Funder
Efficacy and Mechanism Evaluation (EME)	Funder
Chief Scientist Office (CSO)	Funder
National Institute for Social Care and Health Research (NISCHR)	Funder
Medical Research Council (MRC)	Funder
European and Developing Countries Clinical Trials Partnership (EDCTP)	Funder
Horizon 2020	Funder
Medical Council of Canada (MCC)	Funder
Canadian Cancer Trials Group	Sponsor
Experimental and Clinical Research Center (ECRC)	Funder
National Institute of Health (NIH)	Funder
Cancer Research UK	Charitable
Leukaemia UK	Charitable
Lymphoma Research Trust	Charitable
Multiple Sclerosis Society	Charitable
Action Medical Research	Charitable
Age UK	Charitable
Alzheimer's Research UK	Charitable
Arthritis Research UK	Charitable
Asthma UK	Charitable
Breast Cancer Now (formerly Breakthrough Breast Cancer)	Charitable
Breast Cancer Campaign	Charitable
British Heart Foundation	Charitable
Bupa Foundation	Charitable
Cystic Fibrosis Trust	Charitable
Epilepsy Action	Charitable
Epilepsy Research UK	Charitable
Marie Curie Cancer Care	Charitable
Meningitis UK	Charitable
Roy Castle Lung Foundation, The	Charitable
SPARKS - The Children's Medical Research Charity	Charitable
Wellcome Trust, The	Charitable
European Organisation for Research and Treatment for Cancer (EORTC)	Charitable
Association of Medical Research Charities (AMRC)	Charitable
Experimental Cancer Medicine Centres	Funder
IMPACT Clinical Trials Partnership	Funder

Survey of UKCRC Registered CTUs

List of UKCRC Registered CTUs Contacted:

Barts and the London Pragmatic CTU
Barts Clinical Trials Unit
Birmingham Clinical Trials Unit
Bristol Clinical Trials and Evaluation Unit
Bristol Randomised Trials Collaboration
CaCTUS (Cancer Clinical Trials Unit Scotland)
Cambridge Clinical Trials Unit (CCTU)
Cancer Research UK Clinical Trials Unit (CRCTU)
Centre for Healthcare Randomised Trials (CHaRT)
Centre for Trials Research
Comprehensive CTU @ UCL
CR UK & UCL Cancer Trials Centre
Diabetes Trials Unit (Churchill Hospital, Oxford)
Edinburgh Clinical Trials Unit, Edinburgh
Glasgow Clinical Trials Unit
Imperial Clinical Trials Unit
Intensive Care National Audit & Research Centre (ICNARC) CTU
Keele Clinical Trials Unit
King's Clinical Trials Unit at King's Health Partners
Leeds Clinical Trials Research Unit
Leicester Clinical Trials Unit
Liverpool Trials Collaborative
London School of Hygiene & Tropical Medicine
Manchester Clinical Trials Unit
Medical Research Council Clinical Trials Unit at UCL
Newcastle Clinical Trials Unit (NCTU)
NHS Blood and Transplant Clinical Trials Unit
North Wales Organisation for Randomised Trials in Health (NWORTH)
Northern Ireland Clinical Trials Unit
Norwich Clinical Trials Unit
Nottingham Clinical Trials Unit
NPEU Clinical Trials Unit
Oxford Clinical Trial Service Unit & Epidemiological Studies Unit (CTSU)
Oxford Clinical Trials Research Unit (OCTRU)
Oxford Primary Care and Vaccines Collaborative Clinical Trials Unit
Papworth Trials Unit Collaboration
Peninsula Clinical Trials Unit
PRIMENT Clinical Trials Unit at UCL
Royal Marsden Clinical Trials Unit (RM-CTU)
Sheffield Clinical Trials Research Unit
Southampton Clinical Trials Unit
Surrey Clinical Trials Unit
Swansea Trials Unit
Tayside Clinical Trials Unit
The Institute of Cancer Research Clinical Trials & Statistics Unit (ICR- CTSU)
Warwick Clinical Trials Unit
York Trials Unit
Brighton and Sussex Clinical Trials Unit*
Cambridge Epidemiology & Trials Unit*
Derby Clinical Trials Support Unit (DCTSU)*
Exeter Clinical Trials Unit*
Hull Health Trials Unit*
Lancashire Clinical Trials Unit*

**These units had provisional status at the time of survey circulation.*

Appendix 4: Expert Panel

Academic Statisticians (alphabetic by surname)

- Cono Ariti (Centre for Trials Research, University of Cardiff)
- Simon Bond (Cambridge Clinical Trials Unit, Cambridge)
- Sarah Brown (Clinical Trials Research Unit, Leeds Institute of Clinical Trials Research, University of Leeds)
- Michael Cole (Newcastle Clinical Trials Unit, University of Newcastle)
- Carrol Gamble (Liverpool Clinical Trials Centre, University of Liverpool)
- Piers Gaunt (Cancer Research Clinical Trials Unit, University of Birmingham)
- Samantha Hinsley (Cancer Research UK Glasgow Clinical Trials Unit, University of Glasgow)
- Jane Holmes (Oxford Clinical Trials Research Unit, Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford)
- Victoria Homer (Cancer Research Clinical Trials Unit, University of Birmingham)
- Emily Robinson (Royal Marsden Clinical Trials Unit, The Royal Marsden NHS Foundation Trust)
- Matthew Schipper (Departments of Radiation Oncology and Biostatistics, University of Michigan)
- Deborah Stocken (Leeds Institute of Clinical Trials Research, University of Leeds)
- Katrina Walker (Clinical Trials Research Unit, Leeds Institute of Clinical Trials Research, University of Leeds)
- Christopher Weir (Edinburgh Clinical Trials Unit)
- Graham Wheeler (Imperial Clinical Trials Unit, Imperial College London)
- Christina Yap (Clinical Trials and Statistics Unit, The Institute for Cancer Research, London)
- Ly-Mee Yu (Primary Care Clinical Trials Unit, University of Oxford)

Pharmaceutical Statisticians (alphabetic by surname)

- Alun Bedding (Roche Products Ltd.)
- John Kirkpatrick (Roche Products Ltd.)
- Thomas Prior (Janssen Research & Development)

Regulatory Statisticians

- Khadija Rantell (Medicines and Healthcare products Regulatory Agency)

Appendix 5. Explanation and Elaboration of Essential Items

Examples from the given references. Some may have been updated to: remove sensitive information (for unpublished SAPs), negate circular references, and/or to increase compliance within guidelines. Moreover, some examples are fictitious.

Section 1: Administrative Information

Title and trial registration

Item 1a:

Descriptive title that matches the protocol, with SAP either as a forerunner or subtitle, and trial acronym (if applicable).

Explanation:

"The title provides vital information required for trial identification. The title should unambiguously state which trial the SAP relates to and should therefore be identical to the trial protocol with 'Statistical analysis plan' either as a fore runner or sub-title. Ideally the title should identify the study design, population, interventions, and, if applicable, trial acronym." [1]

Example:

"CAMELLIA: A Phase I dose escalation trial of a Humanized Monoclonal Antibody in Haematological Malignancies Statistical Analysis Plan" [2]

"Statistical analysis plan for the BUTEO trial: a single arm, two-stage, multi-centre, phase II clinical trial investigating the safety and activity of the use of a human monoclonal antibody, in the treatment of patients with primary sclerosing cholangitis (PSC)" [3]

Title and trial registration

Item 1b:

Trial registration number.

Explanation:

"A trial registration number should be provided which uniquely identifies a clinical trial and its existence on a publicly-accessible registry. The International Committee of Medical Journal Editors (ICMJE) mandates the registration of clinical trials in a primary register of the World Health Organization (WHO) International Clinical Trials Registry Platform (ICTRP) or in ClinicalTrials.gov before recruitment of the first patient as a condition of consideration for publication [4]. This identifier should be clearly listed in all relevant documentation including the protocol and the SAP." [1]

SAP version

Item 2:

SAP version number with dates.

Explanation:

"Sequentially numbering and dating each SAP version avoids any confusion over which document is the most recent. Transparent tracking of version numbers and amendments facilitates trial conduct, review and oversight. The first final version of a document will be Version 1.0. It is recommended that subsequent final documents will have an increase of "1.0" in the version number (1.0, 2.0, etc.).

While the document is under review, subsequent draft versions will increase by "0.1", e.g., 1.1, 1.2, 1.3, etc. When the revised document is deemed final, the version will increase by "1.0" over the version being revised, e.g., the draft 1.3 will become a final 2.0." [1]

Example:
"Version 1.0 (21 Apr 2017)" [5]

Protocol version

Item 3:
Reference to version of protocol being used.

Explanation:
"Referencing the version of the protocol being used is helpful as it links the SAP to the protocol and serves as a reminder that the SAP is not a standalone document and needs to be read in conjunction with the corresponding version of the protocol. This avoids the need for the author to duplicate information from the protocol in the SAP. If there have been protocol amendments after the SAP has been written then the SAP needs to be reviewed against the amendments, and updated where necessary. The information in SAP revision history table may be extended to record that the SAP has been reviewed in light of protocol amendments but no changes were required." [1]

Example:
"This statistical analysis plan is based on protocol version 5 dated 24 February 2015." [5]

SAP revisions

Item 4a:
SAP revision history.

Item 4b:
Justification for each SAP revision.

Item 4c:
Timing of SAP revisions in relation to interim analyses, etc.

Explanation:
"A clear explanation of the changes made between each version of the SAP is essential, along with a justification for the revision and the date. This is important to maintain transparency. After the first version of the SAP is agreed and signed off, the SAP revision history should include the following information: the previous version number, the SAP section changed, details of the change made along with justification for the revision, and date of revision. A justification for each SAP revision is necessary to document the reasons for changes. This ensures the external validity of the trial as it demonstrates that changes are not being made based on unblinded trial data. From a regulatory perspective when SAP revisions occur after unblinded interim analyses have been conducted the people involved in deciding, writing, or approving the SAP should ideally have no knowledge of unblinded data particularly if the trial will be used for a licence application. In other situations, it may be sufficient for the justification to document the reason for the change is not based upon comparative data and for the approver to have no knowledge of unblinded data." [1]

Example

Table A3a: Version History

Statistical Analysis Plan version:	Reason for update:
Vn1.0, Vd15-Jan-2016	Initial release
Vn2.0, Vd12-Aug-2016	Updated to match protocol: objectives, DLT definition. Using cohorts of 3-5 patients.
Vn3.0, Vd09-Mar-2018	Addition of definitions of populations for analysis and protocol changes

[6]

Roles and responsibility

Item 5:

Names, affiliations, and roles of SAP contributors.

Explanation:

"Individuals who contribute significantly to SAP development should have their contributions described. Listing the SAP contributors, their affiliations and their roles in the SAP development process provides due recognition, accountability, and transparency. Naming of authors and statements of author's contributions is standard for SAPs published in journals such as Trials, but rare in unpublished SAPs. Contributors may be non-signatory members if only the statistician writing the SAP, supervising senior statistician and the chief investigator/clinical lead will sign and approve the SAP." [1]

Signatures of:

Item 6a:

Person writing the SAP.

Explanation:

"The signature of the person writing the SAP is crucial as it identifies who is responsible for the SAP and that they have approved the SAP. In all circumstances this should be signed and dated. If an update has been made then the author of the update should sign the updated version." [1]

Signatures of:

Item 6b:

Senior statistician responsible.

Explanation:

"The signature of the senior statistician responsible for overseeing the trial is important as it highlights that the SAP has been reviewed and approved by an experienced statistician. In some circumstances the senior statistician may be the person writing the SAP and such a dual role should be reflected in the signatories. The signature should always be dated." [1]

Signatures of:

Item 6c:

Chief investigator/clinical lead.

Explanation:

"The signature of the chief investigator/clinical lead demonstrates that they have reviewed and approved the SAP. Once the final version has been approved and signed off it avoids any post-hoc changes being made without the justification and approval of all signatory members to maintain internal and external trial validity. The signature should always be dated." [1]

Section 2: Introduction

Background and rationale

Item 7:

Synopsis of trial background and rationale including a brief description of research question and brief justification for undertaking the trial.

Explanation:

"The full rationale for undertaking the trial and trial background are explained in detail in the protocol so only a brief synopsis is necessary within a SAP to avoid duplication of information. The synopsis should include justification for undertaking the trial, why the trial is needed and description of the research question. This item would be regarded as essential if the SAP is to be accessible externally (e.g., published in a journal or on a website) but is optional if the SAP is an internal document only." [1]

Example:

"There is substantial non-clinical, preclinical and clinical data that the therapy can arrest the autoimmune mediated destruction of pancreatic beta cells by induction of functional Tregs that inhibit islet specific autoreactive Teffs. However, prior to embarking on large proof of concept trials in type 1 diabetes it is essential that the dose of the therapy that induces an increase in Treg proportion while resolving qualitative defects is determined." [7]

"This is a phase I clinical trial of the combination of the experimental drug combination in patients with advanced solid tumours. It is a dose escalation study to establish the recommended phase II dose followed by an expansion phase to further assess tolerability, PK/PD profile and antitumor activity of the recommended dose of the combination." [5]

Objectives

Item 8:

Description of specific question, objectives or hypotheses. It should be made clear what the key objectives are (for example primary and secondary objectives that encompasses toxicity, efficacy, PK, PD, or some combination).

Explanation:

The trial objectives reflect the scientific questions to be answered by the trial, defining its rationale and scope. This information may be provided in sufficient detail within the protocol, in which case a reference would be sufficient. If the protocol contains insufficient detail, then additional detail may be required within the SAP. From the trial objectives or hypotheses, it should be clear whether the final trial conclusions (and where appropriate, the dose to be taken forward), are to be based on

toxicity, efficacy, PK, PD or some combination of the aforementioned. In the scenario where the design is jointly assessing toxicity and efficacy, it should be clear which one is to take precedent in the scenario where they draw different conclusions.

Example:

"ADaPT aims to establish a dose of the treatment sufficient to raise circulating DHEA levels in severely injured trauma and hip fracture patients with rule-based escalation supplemented by Bayesian hierarchical models." [8]

"CLARITY aims to assess the eradication of detectable minimal residual disease (MRD) using the drug combination." [9]

Section 3: Study Methods

Trial design

Item 9a:

Brief description of trial design, including the trial phase and the design method (dose escalation e.g., CRM or single-arm phase II e.g., Simon's Two Stage). If the trial has a randomised element to it, summary information regarding the randomisation, including the allocation ratio, should be specified.

Explanation:

Specify the trial design, including references where appropriate. Including the trial phase (e.g., phase I with dose expansion cohort, or phase I/II) is important in the context of early phase trials as there can be less clear distinction between trial phases. The content and level of detail required in the SAP is directly dependent on the methodology that underpins the trial. By making this apparent at an early stage, it encourages transparency and focuses the SAP.

While randomisation is rare in early phase clinical trials, it can occur. If a trial has a randomised element, it is important to state: i) whether the analysis is intended to be comparative, ii) to provide the allocation ratio, and iii) to specify which aspects are blinded. For example, in the instance of a placebo-controlled trial, the trial may be blind to active treatment vs placebo within cohort, but the dose level used in each cohort may be open.

Example:

"This is a prospective, single centre, cross-sectional, randomised, pharmacokinetics study with rule based escalation supplemented by Bayesian hierarchical models. Further details regarding the proposed Bayesian models can be found in section X. The randomised element of the trial randomises patients 1:1 to receive IMP either orally or sublingually. The randomisation will not formally be comparative but aid the evaluation of the secondary trial objectives. The randomisation will not be blinded." [8]

"This is an open-label, multi-centre, dose-escalating adaptive platform phase Ib/IIa trial. The trial will employ a two-stage modified Time-To-Event Continual Reassessment Method for Partial Ordering (PO TiTE-CRM, described in section X) to determine the Maximum Tolerated Dose (MTD) of the drug in combination with radiotherapy." [10]

Trial design

Item 9b:

Treatment information, including the dose levels of intervention(s). Where appropriate, and if multiple doses are used, the following should also be reported: the ordering and combination (in the instance of multiple agents under investigation) of dose levels, and the dose level to start at.

Explanation:

All relevant treatment information should be made available in the SAP, or suitably referenced to in supporting documents (such as the protocol).

In early phase clinical trials, it may be the case that multiple dose levels of treatment are under investigation. If this is the case, then it is advised that these dose levels and their ordering should be clearly written for all trial designs (not just those with a partial ordering component). Here it may be appropriate to include, or refer to, a trial schema or table which should clearly depict the ordering of dose levels. This removes any potential for confusion or ambiguity. Alternatively, if a dose range or formulation is to be administered (e.g., dose volume = (body surface area x Dose Level)/ Concentration of Drug), this should here be specified.

Similarly, the starting dose level of investigation should be made clear. If the trial is to dose escalate and use small cohort run-ins, this should be stated along with when and how this will stop (i.e., what will trigger the full cohort size to be used).

Where multiple doses are used, it is best practice to specify the dose levels under investigation in advance. However, there are times when this is not possible, such as when instances where IMP is administered by IV. If this is the case, careful and thorough documentation should here be provided regarding how doses will be chosen and dose escalation will occur.

Where only a single dose is under investigation the details provided in this section will be briefer.

Example:

"Five doses ($1x10^{10}$; $3x10^{10}$; $1x10^{11}$; $3x10^{11}$ and $1x10^{12}$) of drug will be investigated using a 3+3 design. The first cohort will be treated at dose $1x10^{10}$. The doses given to subsequent cohorts will be adaptively selected based on the incidence of DLTs. This design will require a maximum sample size of 30 patients, and could stop the trial early if excess toxicity is observed at a dose." [11]

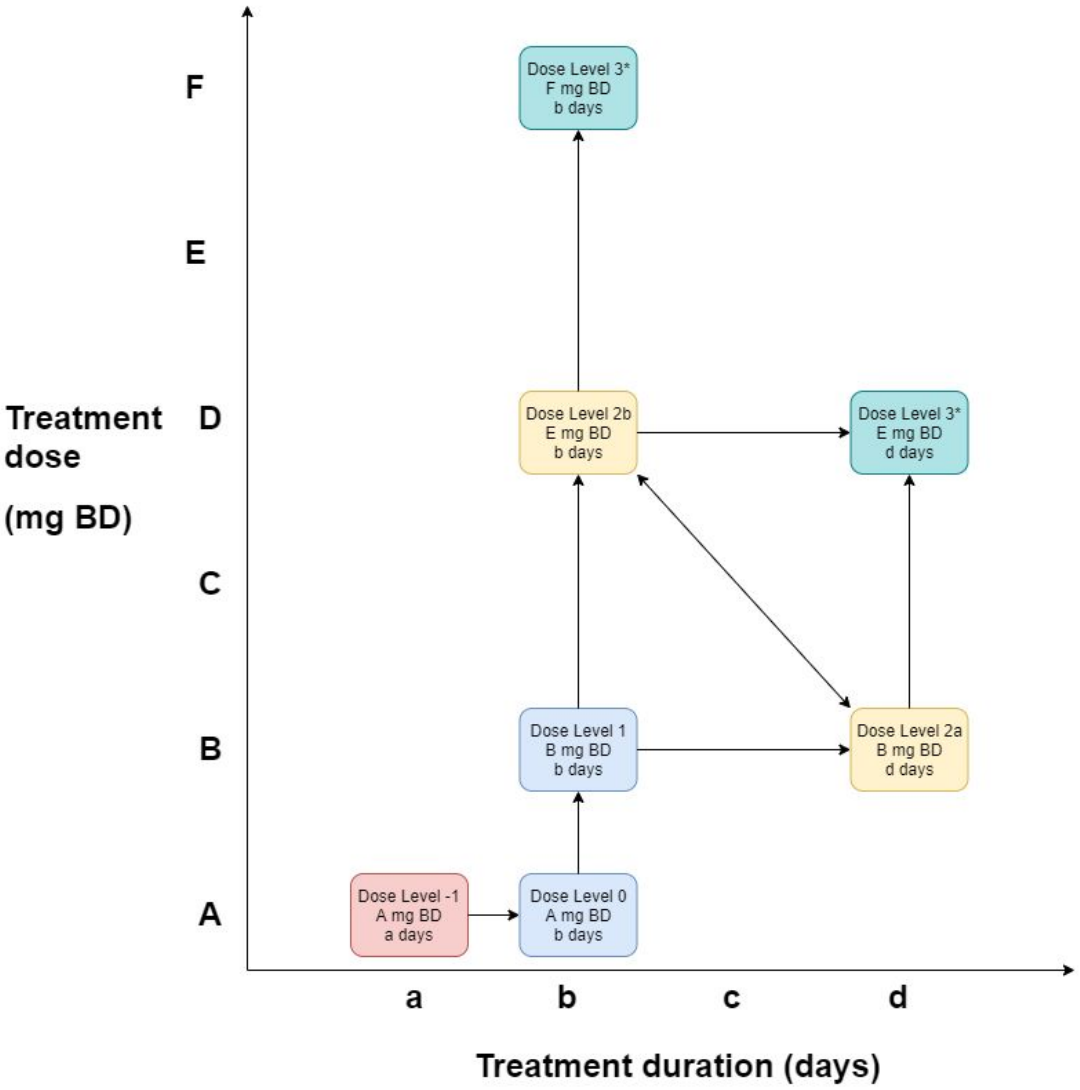


Figure A3a: Dose pathway

"[10]

"Patients will receive x mg/kg of drug subcutaneously for ten weeks once per week in the first instance, during out-patient appointments at site. For the purposes of safety, it is proposed that the first 2 patients will be recruited as Sentinel patients. These will be recruited in series, and each will be assessed for 2 weeks before the next patient will be recruited. Data on Sentinel patients will be assessed by an independent safety monitoring committee. If the safety monitoring committee is satisfied that the product has an acceptable safety profile in the sentinel patients, the study will be opened to general recruitment." [12]

Trial design

Item 9c:

Details regarding the statistical methodology underpinning the trial, including the choice of the number of parameters in the model if applicable, its empirical form and all formulae. It is also important to ensure all model parameters are given, including where appropriate, the weights of the model.

Explanation:

The statistical methodology underpins the trial and ensures that achieving the objectives and hypotheses is feasible. Clear detailing and explanation of the statistical methodology should be made available. This information may be provided within the protocol, in which case a reference to the relevant section(s) of the protocol would be sufficient. However, if the protocol contains insufficient detail, as protocols usually target clinical rather than statistical readers, then additional detail may be required within the SAP. By including details regarding the mathematical form of the model (where appropriate for trial design), and the number of parameters in the model, transparency in the trial is promoted. It may also be appropriate to justify why the model specification was chosen.

Where parameters are to be sampled from a distribution, it is imperative that these distributions, and the elicitation of these (be it through expert elicitation or chosen from standard distributions) are given here to ensure observed results do not influence critical parameters required for analysis.

For models with a TiTE component, the mathematical form of the weight formulae should be explicitly stated.

If the trial makes purely rule-based dose escalation decisions (e.g., 3+3), and there is negligible statistical methodology underpinning the trial design, this section can be omitted.

Example:

"The EffTox design [13, 14] (and version 4.0.12 of the EffTox software, and a proprietary implementation of EffTox written in Python, where necessary) is used for dose escalation/de-escalation decisions. This design establishes the optimal dose which is both safe and effective in terms of the definition of tolerability and efficacy as above.

EffTox estimates the probability of efficacy and toxicity at each dose given the patient outcomes observed and the investigators' prior beliefs. The design then uses contours to calculate the utility score of each dose given its associated probabilities of efficacy and toxicity. A dose is preferable to another if it has a higher utility score. When invoked to provide the next dose allocation, the EffTox design disregards the doses that are probably intolerable or ineffective. Of the remaining doses, it selects the dose with the greatest utility score.

We seek a dose of drug x to be given in combination with treatment y that is associated with a probability of efficacy of 45% or more, and a probability of toxicity of 40% or less. The EffTox design will infer that a dose is probably ineffective if there is at least a 97% probability that the rate of efficacy is less than 45%. It will infer that a dose is probably intolerable if there is at least a 95% probability that the rate of toxicity is greater than 40%.

Initial patients will receive dose level 1. The model will be updated after each patient or cohort of patients is evaluated for DLT and efficacy outcomes. The model is updated using all accumulated information to provide the recommended dose for the next patient. The EffTox design does not skip untried doses in escalation or de-escalation.

When calculating the next dose, EffTox calculates the Bayesian posterior probabilities of toxicity and efficacy at each dose using the patients' outcomes accumulated thus far. Marginal probabilities of toxicity and efficacy are modelled in linear (1) and quadratic (2) form respectively.

The marginal probability of toxicity at dose *x* is given by:

$$\pi_T(x, \boldsymbol{\theta}) = g^{-1}\{\eta_T(x,\boldsymbol{\theta})\} \quad where\ \eta_T(x,\boldsymbol{\theta}) = \mu_T + x\beta_T, \qquad (1)$$

$$with\ \beta_T > 0\ for\ \pi_T(x, \boldsymbol{\theta})\uparrow in\ x\ and\ real\ valued\ otherwise$$

and the marginal probability of efficacy at dose *x* is given by:

$$\pi_E(x, \boldsymbol{\theta}) = g^{-1}\{\eta_E(x,\boldsymbol{\theta})\} \quad where\ \eta_E(x,\boldsymbol{\theta}) = \mu_E + x\beta_{1,E} + x^2\beta_{2,E} \qquad (2)$$

where g^{-1} is the inverse logistic transform. As it is expected that higher doses of the combination do not necessarily result in greater efficacy, a quadratic form is utilised to allow for this non-monotone dose-response relationship.

The joint probability model is:

$$\pi_{a,b} = \pi_E^a(1-\pi_E)^{1-a}\pi_T^b(1-\pi_T)^{1-b} + (-1)^{a+b}\pi_E(1-\pi_E)\,\pi_T(1-\pi_T)\frac{e^{\varphi}-1}{e^{\varphi}+1} \qquad (3)$$

$$where,\ a, b \in \{0, 1\}\ and\ real\ valued\ \varphi\ \ and\ \boldsymbol{\theta} = (\mu_T, \beta_T, \mu_E, \beta_{1,E}, \beta_{2,E}, \varphi)$$

where the x, theta notation has been suppressed for readability and is an association parameter. The model hyperparameters are based on informative prior guesses for efficacy and toxicity at each dose level elicited from clinical investigators (Table A3b).

Table A3b: Prior probabilities of toxicity and efficacy outcomes by dose level.

Dose Level	Dose (mg/m^2) - drug x	Prior Prob(tox)	Prior Prob(eff)
-2	a every other day	0.025	0.20
-1	a once daily	0.05	0.30
1 (starting dose)	b once daily	0.10	0.50
2	c once daily	0.25	0.60

The prior effective sample size (ESS) used is 1.3. Thall *et al.*, [15] advise ESS values between 0.5 and 1.5. The greater the ESS, the stronger weight the investigators' prior beliefs bear on the posterior beliefs. Prior beliefs on the six model parameters are assumed to be described by normal distributions. The hyperparameter values associated with our value for ESS are calculated by the EffTox software (Table A3c).

Table A3c: Hyperparameter prior means and standard deviations.

Parameter	Prior mean	Prior standard deviation
mu_T	-5.4317	2.7643
$beta_T$	3.1761	2.7703
mu_E	-0.8442	1.9786
$beta_{T1}$	1.9857	1.9820
$beta_{T2}$	0	0.2
psi	0	1

"[16]

"For patients who have not completed the scheduled treatments, the TiTE-CRM model will weight their safety data based on the proportion of days the patient has been assessed for over the DLT assessment period using a linear function." [17]

"Linear weighting functions will be employed for any patient with a length of follow up between the three time points. One weight function to calculate weights between 8-12 weeks and another for weights between 12-52 weeks. For the weighting function $w(u;t_1, t_2, t_3)$ where u is the time-to-toxicity of patient j and t_1, t_2, t_3 is the time period with values 8, 12 and 52 respectively. Then for $t_1 \leq u \leq t_3$

$$w(u; t_1, t_2, t_3) = 0.6 + 0.2\frac{min(0, min(u, t_2) - t_1)}{t_2 - t_1} + 0.2\frac{\max(0, u - t_2)}{t_3 - t_2}$$

"[10]

Trial design

Item 9d:

Rules of the trial design and model.

Here information on the target objective (toxicity, response, PK, or PD, either singularly or in combination), classification of overdosing, and any stopping boundaries should be given. This may include the desired certainty in these estimates.

Moreover, where dose decisions (e.g. escalation, de-escalation, remain at current dose or stop early) are to occur, details regarding dose escalation transitions and dose skipping should be given

Explanation:

The primary objectives for most early phase trials are investigated by trying to attain a fixed probability of an event occurring, or value of a continuous outcome. For safety and dose escalation trials, this may be targeting a toxicity probability (e.g., probability of toxicity between 25% and 33%), or attaining a fixed figure from a continuous scale (e.g., an Area Under the Curve 0-24 hours after administration (AUC_{0-24}) above a desired threshold); whereas for single arm phase II trials, this may be targeting an efficacy/response probability (e.g., at least 72% (13/18) of patients achieving an objective response). There do exist trial designs, such as the EffTox and Emax designs, which target both. These targets should be made clear. Moreover, for multi-stage designs where the continuation of the trial is based on formal interim analyses, the target probabilities at each interim, and where appropriate, stopping boundaries (e.g., at the interim if at least 50% of patients have achieved an objective response, then the trial shall continue) should also be explicitly given.

Furthermore, indications of the desired certainty that these targets have been attained should be made distinct. For example, we seek a 70% posterior probability that the true toxicity rate falls between 25% and 33%, or evidence that the lower bound of a 95% confidence interval for protection rate is greater than 70%. For fixed designs, such as the A+B dose escalation, or single arm single stage phase II, such as A'Hern's designs, this certainty is ascertained from exact probability distributions. Where interim analyses and stopping rules are implemented, the desired certainty in the interim results should also be given.

For early phase trials with outcomes or dose escalation decisions that depend on toxicity, classification of over- and under-dosing, and the certainty in these that would warrant action should be given either in the SAP or suitably referenced (for example to the protocol).

Details on how the design would be implemented in the trial and the adaptations that would be made based on accruing data on key outcomes (e.g. toxicity or efficacy or both) should be provided. For instance, for dose escalation trials, explicit statement of the rules regarding dose escalation and de-escalation, especially regarding dose skipping. For example, no doses or only up to one dose may be skipped per escalation, however, doses can be skipped if the dose is to be de-escalated for safety. Instances where the model can be overwritten (due to safety concerns) and the dose selected downgraded require documentation detailing criteria for these situations. This information should either be given in the SAP or suitably referenced (for example to the protocol).

Example:

"We seek a dose of drug x to be given in combination with treatment y that is associated with a probability of efficacy of 45% or more, and a probability of toxicity of 40% or less. The EffTox design will infer that a dose is probably ineffective if there is at least a 97% probability that the rate of efficacy is less than 45%. It will infer that a dose is probably intolerable if there is at least a 95% probability that the rate of toxicity is greater than 40%." [16]

"The Simon's two-stage minimax design requires 3/18 successes at the interim analysis to continue, and 9/37 successes at the final analysis." [3]

"The trial may stop early for safety. In the event that all dose levels are toxic, the trial will stop before reaching the maximum number of patients. If P(risk of DLT > 0.35|dose = 1, current data) > 0.65 for the lowest dose level and at least three patients have complete data for the toxicity endpoint (a DLT or have completed the toxicity window) we will stop the trial.

The dose suggested for the next patient is the optimal dose as defined above. However, escalation to an untried dose is subject to no dose skipping, and is only permissible if at least 2 patients have been given the dose immediately below for at least 8 weeks. There is no restriction on de-escalation." [18]

Trial design

Item 9e:

Experimental details and design specifics.

For dose escalation trials, information regarding cohort size, including whether this is fixed or flexible should be given.

Indication of the stopping rules for interim and final evaluations.

For model-based and model-assisted designs, details on the prior including full skeleton (if applicable) and its elicitation should be given.

For single arm phase II trials, the target sample size and, where appropriate, the timing of any interim analyses.

Explanation:

Most dose escalation trial designs rely on patients being enrolled in dose cohorts and then being evaluated at the appropriate time before decisions regarding dose escalation are made. Therefore, it is imperative to include information regarding cohort size, the target total trial sample size, and the timing of dose escalation decisions. It should also be made clear whether the total cohort size is going to reflect when the dose escalation decisions will be made. For example, if the total cohort size is going to be n=8, but dose escalation decisions made when n>4, or after each patient in the cohort

has at least completed cycle 4. Sufficient detailed regarding cohorts and dose escalation decisions may be contained within the protocol, in which case suitable reference to this section is sufficient. If flexible cohort sizes are to be used, then reference should be made to how the size of the cohort will be ascertained.

For novel agents, sentinel dosing may be used to aid safety evaluation before the recruitment to the full cohort commences. If sentinel participants are to be used, a clear description of the number of sentinel participants, how much treatment and follow up they need to complete before the recruitment to the full cohort can commence, and whether they will be evaluated with the full cohort should here be included. If adequate details are captured in a supporting document (such as the protocol), suitable reference to these may instead be given here.

Where model-based or model-assisted designs are used and continually updated, thus where the notion of cohorts is depreciated, information regarding when the model will be updated should be specified (e.g., after each DLT or at least every three evaluable patients or a minimum treatment period).

Where simulations have been run to assess the operating characteristic of the trial design, summary details may be given here with reference given to another document (such as a simulation report) where greater details are contained.

Dose expansion cohorts are often used to gain a better insight into the safety or efficacy profile at the proposed dose if these are to be used, information regarding the sample size of the expansion cohort should be included here. Moreover, if results from the dose expansion cohort contradict those from the original dose escalation trial, clarification should be provided regarding the consequences (e.g., if doses could be altered).

For single arm phase II trials reference to the total and, where appropriate, interim sample sizes should be made (e.g., the interim analyses will take place after the first 9 evaluable patients have received their outcome assessment visit). It is not necessary to include the full power and sample size calculations, as these will be detailed later (see item 11).

A definition or suitable reference to, the end of trial definition, including any formal stopping rules should be included. If simulations have been run to assess the operating characteristics of the stopping rules, either inclusion or reference (for example to a simulation report) to these should be made.

For model-based and model-assisted designs, since the choice of prior distribution, and where appropriate the skeleton distribution, can influence the posterior results, transparency regarding model specification is encouraged and so the full form of the priors should be included in the SAP or be suitably referenced. This further allows for full trial reproducibility and replication if needed. Moreover, an indication as to how this was elicited (e.g., using an expert or expert panel, or using statistical packages, functions or programs) should also be included for transparency.

Example:

"30 patients will be recruited. Simulations have been used to justify this sample size, results are given within the simulation report in Appendix X.
The trial may stop early for safety. In the event that all dose levels are toxic, the trial will stop before reaching the maximum number of patients. If P(risk of DLT > 0.35|dose = 1, current data) > 0.65 for the lowest dose level and at least three patients have complete data for the toxicity endpoint (a DLT or have completed the toxicity window) we will stop the trial." [18]

"There will be a maximum of 12 patients treated in each group in the phase I component. Once a dose has been decided upon for each group there will be a 9 patient expansion in each of these doses for phase II.

Table A3d: Patient Cohorts

Group A (Adult)	Phase I: 12 patients	Phase II: 13 (9 + 4) patients
Group B (Paediatrics)	Phase I: 12 patients	Phase II: 13 (9 + 4) patients
Total for Phase I: max 24	Total for Phase II: max 18	Total for Trial: max 42

"[19]

"The phase II Simon's two-stage minimax design incorporates an interim analysis of the accumulating data. The interim analysis (stage-1) takes place once 18 patients have been evaluated for the primary outcome – which is based on the response level of ALP. If three or more successful responses (i.e. 25% or more reduction in ALP level) are observed in stage-1 then the trial will continue into stage-2. Recruitment will not be halted while stage-1 is assessed. Further patients will be recruited during stage-2 in order to obtain the necessary sample size of 37 patients; allowing for 10% patient drop out during trial duration, this number could reach a total of 41 patients being required." [3]

"The six dose levels scheduled for a combination of drug x and drug y, together with the prior probabilities of a DLT at those levels, are presented in Table A3e. The prior guess of MTD is at Dose 4, but to exercise caution as this combination regimen has never been studied in this patient population, we will start at Dose 2. If the combination dose is too toxic, the design allows for de-escalation to dose level 1.

Table A3e: Dose levels with initial estimates of probabilities of DLT at each level

Dose Level	Drug x dose	Drug y dose	Prior probability of DLT
1	m mg/m^2	a/A mg/m^2	0.05
2 (starting dose)	n mg/m^2	a/A mg/m^2	0.10
3	n mg/m^2	a/B mg/m^2	0.15
4 (prior estimate of MTD)	o mg/m^2	a/B mg/m^2	0.25
5	o mg/m^2	a/C mg/m^2	0.35
6	p mg/m^2	a/C mg/m^2	0.50

"[6]

Randomisation

Item 10:

Where appropriate, randomisation details e.g., whether any minimisation or stratification occurred (including stratifying factors used or the location of that information if it is not held within the SAP) and where applicable, details on blinding.

Explanation:

While randomisation in the context of early phase clinical trials is uncommon, it can occur. If randomisation is used, this should be clearly stated and details regarding the randomisation provided. This will typically include the method of randomisation, e.g., stratification, block, or minimisation and information of factor levels provided (where appropriate). It may be that sufficient information is available in other trial specific documents (such as the protocol), in which case reference to this is acceptable.

Example:

"Approximately 36 eligible subjects aged 10-17 years were to be randomised at a ratio of 1:1:1 to one of three doses of Ferric Maltol (7.8 mg, 16.6 mg or 30 mg BID) for nine days (Days 1 to 9). Randomisation was to be stratified by age (10-14 years, 15-17 years) and gender (male, female)." [20]

Sample size

Item 11:

Full sample size calculation determination or justification or reference to relevant sample size calculation section in protocol (instead of replication in SAP)

Explanation:

The sample size calculation may be included in full in the SAP or a reference to the sample size calculation in the protocol or other document may be provided. The sample size calculation is an important piece of information for every trial as it determines how many patients are required in the primary analysis to ensure the trial is appropriately justified to detect a clinically important difference.

For phase I trials, it may be sufficient to justify the trials sample size by the number of patients per cohort, and the total number of cohorts expected to be enrolled. Moreover, for dose escalation trials where dose escalation is based solely on the observed toxicity, it may be useful to detail the minimum number of participants expected to be recruited in the scenario that either no DLTs are seen (if this is different to the maximum sample size), or all doses are found to be too toxic.

Where the sample size has been verified by simulations (to ensure the trial can yield a successful result), the operating characteristics and results should be included. Again, it may be appropriate to only include summary information in this section of the SAP with suitable reference given to a supporting document such as a simulation report (see point 33) where greater detail is given.

Where single-arm phase II trial designs are used, it is important to include all relevant information on which the trial design is based, e.g., design (A'Hern, Simon's Two Stage, etc.), statistical significance level (alpha), power, the exact type I and type II error rates (where calculated), effect size including p_0 (the largest unacceptable response rate) and p_1 (the smallest acceptable response rate), and where appropriate, the dropout rate assumed. Moreover, where fixed designs are to be used, it is important to clearly document how deviations from the planned sample size will affect decisions regarding the conclusions drawn from the trial. For example, if a trial requires 22 responses out of 30 patients to be classified as a success, how an increase or decrease in the number of evaluable patients, (e.g., to 32 or 27), would affect the number of required responses and success criteria.

In all scenarios, details of any sample size calculations, including the software used (and version), must be provided to allow the calculation to be reproduced.

Example:

"There was no formal sample size calculation in the Phase I stage. The design was based on the traditional 3+3 design for phase I trials. The recruitment plan was to recruit 3-6 participants to be treated at each of up to 7 dose levels until the MTD could be identified. Participants who did not complete the first cycle of treatment for reasons other than toxicity were replaced at the current dose level. A maximum of 42 participants evaluable for toxicity were required to complete all of the dose levels for Phase I." [21]

"We use an A'Hern design to investigate whether 12 months of combined treatment of drug x and drug y leads to MRD eradication in the bone marrow of at least 30% of patients. Over this time horizon, using drug x as a monotherapy, we would expect no more than 10% of patients to eradicate MRD from their bone marrow, thus we compare p_1=0.3 to p_0=0.1. Using statistical significance (alpha) of 2.5% and statistical power of 95.5%, this design requires at least 10 patients to achieve MRD-eradication in the bone marrow out of 50 to approve the combined treatment.
This means:

- If the true rate of MRD-eradication in bone marrow after 12 months of treatment with drug x & drug y is 10%, the statistical design will correctly reject the treatment at least 97.5% of the time;
- If the true rate of MRD-eradication in bone marrow after 12 months of treatment with drug x & drug y is 30%, the statistical design will correctly approve the treatment at least 95.5% of the time."[9]

"The Simon's two-stage design requires a total of 37 evaluable patients receiving the confirmed dose.
This was calculated using the following parameters: alpha = 0.10, beta = 0.2, p_0 = 0.15, p_1= 0.30.
[**Note**: the values for p_0 (0.15) and p_1 (0.30) correspond to the required reduction in patients experiencing raised levels of ALP from 85% to 70%, i.e., 1-0.85=0.15 and 1-0.70=0.30]
The Simon's two-stage minimax design requires 3/18 successes at the interim analysis to continue, and 9/37 successes at the final analysis.
The phase II Simon's two-stage design incorporates an interim analysis of the accumulating data. The interim analysis (stage-1) takes place once 18 patients have been evaluated for the primary outcome – which is based on the response level of ALP. If three or more successful responses (i.e., 25% or more reduction in ALP level) are observed in stage-1 then the trial will continue into stage-2. Recruitment will not be halted while stage-1 is assessed. Further patients will be recruited during stage-2 in order to obtain the necessary sample size of 37 patients; allowing for 10% patient drop out during trial duration, this number could reach a total of 41 patients being required.
If overall there are nine or more successful responses from 37 evaluable patients, then we conclude that the treatment warrants further investigation. If the prescribed patient number is not met then the appropriate decision criterion, corresponding to the total number of evaluable patients, will be selected from table A3f. Patients treated at the confirmed dose during the dose confirmatory stage will contribute to the total evaluable patient requirement.

Method for calculation used "Sample Size Tables for Clinical Studies Software", Sze-Huey Tan (2008).

Table A3f: Possible stop/go guidelines and associated error rates for the Simon's Two-Stage design utilised for the Phase II trial component (bold indicates design parameters chosen)

Patients (stage 1 interim)	**Responses (stage 1 interim)**	**Patients (total)**	**Responses (total)**	**Type-1 Error (α)**	**Power (1-β)**
18	3	34	8	0.120	0.826
18	3	35	8	0.134	0.846
18	3	36	8	0.149	0.863
18	**3**	**37**	**9**	**0.087**	**0.806**
18	3	38	9	0.099	0.827
18	3	39	9	0.111	0.846
18	3	40	9	0.124	0.861
18	3	41	10	0.073	0.809

"[3]

Framework

Item 12:

If applicable, specify whether trial is to be performed under hypothesis testing or Bayesian framework.

Explanation:

This section is not always appropriate. Relevant details on phase I trials will typically be captured elsewhere.

For single arm phase II trials where hypothesis testing is to be undertaken, outline the intended analysis framework.

Regardless of the framework of the primary analysis, other estimands may be important to draw trial conclusions, for all early phase trials. The SAP should clearly specify the framework for each estimand or provide a global statement.

Example:

"The main analysis for the single-arm cohorts will be Bayesian in nature." [22]

"The A'Hern's design is employed under a frequentist hypothesis framework."

Statistical interim analyses and stopping guidance

Item 13a:

Information pertaining to interim dose decisions (e.g. escalation, de-escalation, remain at current dose or stop early).

Explanation:

Dose-escalation can be poorly documented meaning dose-escalation decisions may be ambiguous. Clear descriptions of the dose-escalation procedure and associated analyses should be provided. This will typically include who will perform the analyses, what interim analyses will be carried out, when they will be performed (e.g., timing and frequency), and who will ultimately decide whether to escalate the dose (e.g. the model, or the DMC/TSC/TMG). Clearly documenting the timing of dose escalation decision in relation to data collection and portion of trial lapsed avoids dose escalation decisions being made based on non-robust/incorrect data. If the trial does not have a dose escalation portion, this section is not necessary. If separate SAPs have been written for dose escalation analyses, then these should be referenced.

Example:

"After the initial 10 patients are assigned to fixed doses as described in Section X the data will be examined by the Dose Determination Committee (DDC) following each new patient. A set of interim analyses will be conducted where the accumulated data will be analysed. A set of candidate models (presented in Section Y) will be fitted to the data. Each model will provide an estimate and standard error (SE) of the target doses that achieve the two targets of a minimum Treg increase and a therapeutic Treg increase. Each model will also provide a recommended dose to assign to the next patient. At each DDC meeting, the choice of the next dose to assign to the next patient will be decided. The choice will be made after consideration of the analyses, but will not be bound by formal decision rules. The choice of dose will always lie below the maximum of 1.5 X 10^6 IU/m^2 BSA.

The target response rates are those that achieve a:

1. Minimum Treg increase defined by the Trial Steering Committee (TSC) at a 10% maximum increase of Treg
2. Therapeutic Treg increase, defined by TSC at a 20% maximum increase" [7]

"The recommended dose (the dose with estimated DLT probability closest to the target of 35%) for each of the subsequent cohorts is determined using the CRM incorporating all of the accumulated DLT outcomes but for added safety, the design includes a restriction to prevent skipping of untested doses when escalating. Recruitment continues until either the maximum sample size is reached, the trial is stopped early due to unacceptable levels of DLT at the lowest dose or when there are four consecutive cohorts allocated to the recommended MTD (providing sufficient evidence that the MTD is reached). The two early stopping rules allow for early termination:

1. If there is a high probability (> 0.7) that the posterior probability of DLT at the lowest dose is greater than the target DLT rate of 35%, indicating that the lowest dose is too toxic.
2. If four consecutive cohorts (three patients in each cohort) have already been allocated at the current MTD, which would also be the recommended dose level for the next cohort if the trial continued.

The value of 0.7 was selected so that the design will recommend stopping early for excessive toxicity if we observe 2 or 3 DLTs out of the first 3 patients at the lowest dose level." [23]

Statistical interim analyses and stopping guidance

Item 13b:

Information on other interim analyses specifying what and when interim analyses will be conducted.

Explanation:

Information needed to conduct any other interim analyses, aside from dose-escalation analyses, should be detailed. Information to be recorded includes statistical methods to be used, who will

perform the analyses, what interim analyses will be carried out, when they will be performed (e.g., timing and frequency), and what decisions can be taken. If there are multiple interim analysis time-points, researchers may choose to include checklists detailing which analyses are to be carried out at each time point. If interim analyses are not planned then this should be stated for clarity. Moreover, if sample size re-estimations to verify initial assumptions are to be performed following such interim analyses, indication of this and the assumptions which are liable to be tested (e.g. variance of the primary outcome, overall event rates, dropout rates) should be here detailed. If details of interim analyses are recorded with sufficient detail in other documents, such as the protocol, then suitable reference may be appropriate to avoid duplication. Finally, if separate SAPs have been written for interim analyses, then these should be referenced.

Example:

"Only one interim analysis is planned and will take place once 18 patients have been evaluated for the primary outcome (response in ALP level, measured from baseline to day 99). The interim report will be prepared and supplied to the DMC when the study has recruited and evaluated 18 patients at the chosen MED dose of drug X (including those recruited on the MED dose during the dose confirmatory stage), or annually whichever is earliest." [3]

"No formal interim analysis is planned for this trial. However, accumulating un-blinded data will be presented by component/cohort and treatment arm on a yearly basis to an independent Data Monitoring Committee (DMC) for monitoring of safety, recruitment, data quality and activity. After the trial has opened, a Trial Safety Committee (TSC), with an independent chair, will meet at least annually following the DMC to provide overall supervision for the trial and provide advice through its independent chair. The ultimate decision for the continuation of the trial lies with the TSC." [22]

Statistical interim analyses and stopping guidance

Item 13c:

Any planned adjustment of the significance level due to interim analysis.

Explanation:

Many early phase trial designs feature formal interim analyses, both in the context of dose escalation trials or multi-stage single arm phase II trial designs, to inform the future conduct of the trial. These interim analyses and where appropriate, any adjustments to control the type I error rate, are often informed by the trial design. If alpha spending functions are going to be used to control the type I error rate, the chosen approach should be clearly specified, justified and referenced. If no adjustments for alpha spending are to be made, this should also be clearly stated.

Example:

"This is not a confirmatory study, we will not consider multiple testing although we do acknowledge that any finding relating to secondary endpoints will be treated as hypothesis generating." [7]

"There are three sources of multiplicity in this study: multiplicity due to interim analyses, multiplicity due to multiple doses, and multiplicity due to multiple endpoints.

The overall type 1 error rate for the study is protected against multiplicity due to interim analyses, because of the alpha- and beta-spending rules described in the preceding section.

The overall type 1 error rate for the study will be protected against multiplicity due to multiple doses by using a step-down, or gatekeeper, procedure. The statistical significance of the difference in response between the low dose group and the placebo group will be assessed if and only if the

corresponding difference between the high dose and placebo has already been shown to be statistically significant.

The study has a single primary endpoint, corresponding to the single primary estimand.

All other endpoints are secondary or exploratory. Therefore, no adjustment to nominal p-values will be made to protect the overall type 1 error rate for the study against multiplicity of endpoints." [24]

Statistical interim analyses and stopping guidance

Item 13d:

Details of guidelines for stopping the trial early.

Explanation:

Details should be provided on the guidelines to be used for stopping the trial early, including whether these stopping rules are binding or advisory and any alternations to recruitment which may be implemented prior to stopping the trial early.

Information on specific stopping boundaries and/or thresholds to be used, including posterior probability cut-offs should be included.

A description of instances where model prediction can be overridden for safety reasons should be pre-specified. The risk of overdosing should be quantified and justified during the design. Such calculations will often be given in supporting documents, e.g., in the simulation report or the protocol. Reference to these documents should be made. It should be clear whether a statistical method will be considered within the early stopping guidelines.

Example:

"Two additional criteria have been added to the modified TiTE-CRM to allow for early termination of either group. They are as follows:

- If there is a high probability (>80%) that the posterior probability of DLT at the lowest dose is greater than the target DLT rate, indicating that the lowest dose is too toxic. If the model recommends early stopping due to this safety criteria, the TMG and Safety Committee will be alerted and the latter, with support of any external evidence, will recommend if the trial should be stopped.
- We would allow the trial to stop early before the full recruitment of 21 patients if nine patients have already been allocated at the most current MTD, which would be the recommended dose level for the next cohort if the trial continues, in consultation with the DMC.

A "look ahead" strategy will be implemented if the next recommended dose level by the modified TiTE-CRM model will not be influenced by the outcome of the remaining patient(s) of a particular cohort (DLT or no DLT). By implementing this strategy, we enable the next cohort of patients to be recruited immediately without awaiting the final observations from the current cohort, therefore reducing waiting times." [17]

"The clinical trial will be subject to periodic reviews by an independent safety monitoring committee. The trial will be suspended if any of the following conditions are met:

1. ≥1 patient in the first Sentinel patients experiences a Serious Adverse Event related to IMP

2. ≥33% of patients (with n > 3) recruited to the study show a significant decrease in the functional rating score (>50%) compared to baseline during the 10-week dosing period.
3. ≥33% of patients (with n > 3) recruited to the study show a significant decrease in Quality of Life (>50%) compared to baseline during the 10 week dosing period." [12]

Timing of final analysis

Item 14:

Timing of final analysis, e.g., all outcomes analysed collectively or timing stratified by planned length of follow-up.

Explanation:

"Information on the timing of final analyses should be included, if relevant. Information on timing of final analysis should explain whether all outcomes are analysed collectively or whether timing is stratified by length of follow-up required. Details should be provided on whether there are short-term and long-term outcomes and how they will be reported i.e. will all outcomes be analysed collectively or will the short-term outcomes be published earlier and the long-term outcomes reported at a later date." [1]

Example:

"A preliminary final analysis will be undertaken to present available data for the escalation phase of the study once an MTD has been determined and agreed. Once the escalation phase is complete and all patients have been followed up for the full duration in accordance with the schedule then the planned final analysis for this drug will be undertaken. This will take into account secondary and exploratory outcome measures." [10]

"For this study, the end of the trial is defined as "the final visit of the last patient undergoing the trial". A final visit should take place 30-35 days after the last administration of IMP. All patients will be followed up for survival (unless they withdraw their consent) once every 3 months until death or until the last patient last visit (LPLV) time point, whichever occurs first. After the LPLV, the trial data will be monitored, then locked, final data listings will be produced and the analyses will be carried out." [2]

Timing of outcome assessments

Item 15:

Time points at which the outcomes are measured including visit "windows".

Explanation:

"The time points at which outcomes are measured is helpful information that can be found in the protocol often in table format. The SAP should either refer to the relevant section of the protocol for details or include this information. If outcomes are required to be measured within a particular time window in relation to each planned visit in order to contribute to the analysis then this should also be specified." [1]

Example

Table A3g: Endpoints & Timing of Evaluation

Endpoints/ Outcome measures	**Time point(s) of evaluation of this end point**
• Dose limiting toxicity • MRI tumour regression grade	• Daily from Day 1 to Week 13 • Week 13
Endpoints	
• Treatment tolerance measured by proportion of patients completing at least 80% of the intended Capecitabine dose and at least 20 fractions of radiotherapy	• Week 9
• MRI tumour regression grade • Pathological complete response • Neoadjuvant Rectal (NAR) score	• Week 13 • Post resection • Post resection

[18]

Section 4: Statistical Principles

Indications of uncertainty [a]

Item 16:

Level of statistical significance.

Explanation:

Where applicable and if traditional tests of significance and cut-off values are to be used to gauge statistical significance, then the significance level to be used including whether tests will be one- or two-sided should be documented. Where a trial has a formal sample size calculation, the significance level used for the primary outcome should be consistent with that used in the sample size calculation. However, it is not necessary for secondary outcomes to use the same significance level, and if these are to change depending on outcome, the critical value for each outcome should be documented.

Example:

“There is no statistical significance level defined for the primary outcome in CAMELLIA as it is a dose-finding trial and does not involve hypothesis testing; there will be no adjustment for multiplicity. The secondary outcomes described here will be assessed at the 5% significance level and/or using 95% two-sided confidence intervals, as appropriate.” [2]

“Unless specified otherwise, a two-sided significance level of 5% will be used in frequentist analyses.” [8]

Indications of uncertainty [a]

Item 17:

Description of any planned adjustment for multiplicity, and if so, including how the type I error is to be controlled.

Explanation:

Multiple testing in the context of early phase trials is generally not recommended, as these trials tend not to formally test hypotheses, rather make recommendations for future confirmatory trials. [25] However, if adjustments for multiplicity are to be made, authors should pre-define what methods will be used and which outcomes these methods will be applied to. The rationale for adjustment and method(s) chosen should also be justified.

Example:

"There will be no adjustment for multiplicity in this trial." [21]

Indications of uncertainty [a]

Item 18:

Either confidence or credible intervals to be reported (appropriately picked dependent on the trial methodology).

Explanation:

The intervals (either confidence or credible) are essential to the interpretation of statistical analyses reported for any of the primary or secondary outcomes. Typically, confidence intervals (CI) and p-values will be reported if the trial uses a frequentist framework, whereas credible intervals (CrI), and where appropriate posterior probabilities, will be reported if the trial uses a Bayesian framework. The level of the CI or CrI to be reported should be decided at the design stage to avoid bias being introduced by modification based on trial data. These levels may be consistent across outcomes or vary by primary, secondary, exploratory and safety outcomes. If this is the case, this should be clearly specified.

If models are being implored at any point, it may be appropriate to here specify the model output which will be reported.

Example:

"95% confidence intervals, calculated using Wilson's method, will be used in frequentist analyses." [8]

"The proposed target doses of each model with their standard errors and with a 95% confidence interval will be reported." [7]

"The posterior probability of DLT at each dose level will be reported with 95% credible intervals."

Adherence and protocol deviations

Item 19a:

Definition of adherence to the intervention and how this is assessed including extent of exposure

Explanation:

"Authors should pre-specify their definition of adherence to the intervention. Non-adherence to the intervention can include not completing the intervention, (e.g., not consuming all prescribed drugs

or consuming a lower dose than is prescribed). This may be reported to aid generalizability of results or may be linked to an analysis population specification." [1]

Example:
"Adherence/Compliance will be assessed by the date of protocol treatment, dose delays, discontinuation and reasons for delays or discontinuation for each patient." [21]

Adherence and protocol deviations

Item 19b:
Description of how adherence to the intervention will be presented.

Explanation:
"Along with defining adherence to the intervention it is also crucial to describe how adherence to the intervention will be presented. This process avoids any bias being caused by adherence being defined after unblinding of data." [1]

Example:
"The treatment that patients received in each cohort will be reported in table X (patient disposition and treatment) and figure Y (treatment received by cohort). Specifically, the treatment received, dose delays, dose intensity, discontinuation and reasons for delays or discontinuation will be reported." [21]

Adherence and protocol deviations

Item 19c:
Definition of protocol deviations for the trial.

Explanation:
"A protocol deviation is defined as a failure to adhere to the protocol such as the wrong intervention being administered, incorrect data being collected and documented, errors in applying inclusion/exclusion criteria or missed follow-up visits. A protocol deviation should be defined as major or minor. A deviation may be considered a serious breach if it affects efficacy, the safety, physical or mental integrity of the participants in the trial, or the scientific value of the trial. Protocol deviations should be defined prior to unblinding of data to avoid any bias being caused and due consideration given to inclusion of participants within analysis populations. [26] Protocol deviations may be defined in another document and referenced within the SAP." [1]

Example:
"A protocol deviation is defined as a failure to adhere to the protocol. Major and minor deviations are defined in the protocol. A deviation may be considered a serious breach if it affects efficacy, the safety, physical or mental integrity of the participants in the trial, or the scientific value of the trial. For this study protocol deviations will be defined as deviations from the treatment schedule as per the protocol." [21]

Adherence and protocol deviations

Item 19d:
Description of which protocol deviations will be summarized.

Explanation:

"A description should be provided on how protocol deviations will be summarised. Providing details of whether the deviation is major or minor is helpful if sensitivity analyses are to be conducted by removing patients with major deviations to assess impact on overall conclusions or to align with analysis populations. The approach to summarising the protocol deviation should also be made clear e.g., number and type of protocol deviations by intervention group or listing of all deviations." [1]

Example:

Protocol deviations will be reported for each dose level, tabulated according to their major/minor classification.

Analysis populations

Item 20:

Clear definition of the trial/dose cohort(s) including how cohorts will be referred to, how patients enter cohorts, the minimum number of patients needed to be in a cohort (and how long they have been in) before dose escalation decisions can be made.

Trial level definitions of patient populations (e.g., per-protocol, intention to treat, safety) should also be given.

Details regarding evaluable patients and specify what happens to unevaluable patients should also be made.

These definitions should be also be provided for any interim analysis populations.

Explanation:

The analysis populations should be specified in advance. This includes how the analysis populations will be defined and which dose escalation decisions and outcomes will be analysed according to each analysis population. It is important to clearly define populations, even if terms are considered standard. For example, if there is no consistent definition of intention to treat (ITT) and the phrase has different meanings for different authors, then a clear definition of these patient populations facilitates the definition of outcomes under the estimands framework (further details given in section 6: Analysis). Patients may be evaluable for different populations.

In the context of dose escalation trials, it is also important to define the cohorts and how they will be referred to, (for example, according to the dose they received or their sequential enrolment). It should also be made clear how many patients can enter each cohort, and the minimum number per cohort and how much trial treatment/follow-up they must have completed before dose escalation decisions can be made.

For all types of early phase trials, the criteria for a patient to be considered evaluable for outcome assessment and when patients are to be replaced should be stated (e.g., must complete at least one IMP administration). It is common in early phase trials, that patients who are not evaluable (for example due to withdrawal or non-compliance) are replaced.

In the event that the trial has a formal sample size calculation and does not recruit to target, it should be specified what the minimum percentage of the target sample size that would need to be recruited to justify completing the full analysis. For recruitment below this threshold, it should be detailed what analysis will be performed and reported.

This section should be made clear either in the SAP, or suitably referenced supporting document (e.g., trial protocol).

Example:

"Dose escalation population: Assessment of the proportion of DLTs for each dose level will be based upon assessment of patients who complete ≥75% of their doses (≥ 6 doses) during the DLT assessment period or who experience a DLT at any time after initiation of the infusion of the first dose. Patients who withdraw early from the study for reasons other than DLT will not be assessable for DLT, and may be replaced by another patient at the same dose level.
Safety population: The safety analysis population will include all patients who received at least part of one dose of drug X. Efficacy (disease response) population: The efficacy population will include all patients who have received at least part of one dose of study treatment and at least one post-treatment response assessment." [2]

"All patients will be analysed on an intention to treat basis. Any patients discovered to be ineligible after being entered into the trial will be listed. Participants who did not complete the first cycle of treatment for reasons other than toxicity were replaced at the current dose level. A maximum of 42 participants evaluable for toxicity were required to complete all of the dose levels for Phase I. All patients starting cycle 1 treatment were evaluable for toxicity." [21]

"The primary analysis population will consist of all participants who receive at least one dose of any trial treatment and, have at least one response assessment available. Only participants, for whom written informed consent has not been received, will not be included in this population.

The safety population will include all participants who receive at least one dose of any trial treatment. Only participants for whom written informed consent has not been received, will not be included in this population." [27]

Section 5: Trial Population

Screening data

Item 21:

Reporting of screening data (if collected) to describe representativeness of trial sample.

Explanation:

"If a trial collects screening data then it is important that the data are appropriately presented to describe the representativeness of the trial sample. This information is not only important for the trial but also important for future trials in the area. The process for screening patients e.g. how patients will be screened and what data will be collected, should be fully described within the trial protocol. According to the CONSORT guidelines [28] as a minimum the number of patients who are assessed for eligibility should be provided with this information presented in a flow diagram, however, more detailed tabulations may be provided. The SAP should describe how this data will be summarised and presented." [1]

Example:

"Information relating to screening data including the number of participants screened, found to be ineligible (with reasons where available) or declined to participate (with reasons where available) will be presented as in Table X." [2]

Eligibility

Item 22:

Summary of eligibility criteria.

Explanation:

"The trial inclusion and exclusion criteria should be specified in the protocol. Details of how eligibility data will be summarised should be provided. Some CONSORT diagrams provide details of the number of patients screened followed by a breakdown of how many patients were eligible and how many were excluded due to violating each inclusion/exclusion criteria." [1]

Example:

"The number of patients falling into the exclusion criteria will be tabulated by cohort and any ineligible patients randomized will be reported, with reasons for ineligibility in Table X." [21]

Recruitment

Item 23:

Information to be included in the CONSORT flow diagram.

Explanation:

"Information included within a CONSORT flow diagram displays the progress of all participants through the trial. The CONSORT guidelines say that "you must complete a flow diagram in order to be compliant with the CONSORT 2010 standard." [28] They provide a CONSORT flow diagram template that can be used and adapted to create a trial specific flow diagram. All necessary information that is displayed in a CONSORT flow diagram should be listed in the SAP so it is clear where the patient throughput will begin to be summarised and how, specific follow-up time points that will be presented along with information on withdrawals and loss to follow up. Alternatively, a study specific CONSORT flow diagram template can be included in the SAP highlighting the information that will be collected." [1]

Example:

"The flow of participants through each stage of the trial, including numbers of participants assigned to a schedule, receiving intended treatment, completing the study protocol, and analysed for the primary outcome is provided following CONSORT. Protocol violations/deviations and information relating to the screening data including the number of ineligible patients entering the study, together with reasons will be reported. Information on number of participants screened, found to be ineligible (with reasons where available), refused to participate (with reasons where available) will also be included.
A CONSORT diagram will be prepared, an example CONSORT diagram is given in Appendix 3." [18]

A CONSORT diagram will be produced to highlight the flow of patients through the trial, and a dose decision by cohort diagram will be produced to show the number of patients enrolled to each cohort and the decisions of the DDC.

Withdrawal/follow-up

Item 24a:

Level of withdrawal, e.g., from intervention and/or from follow-up.

Explanation:

"In this section, all the possible levels of withdrawal should be listed, which may differ from trial to trial. Participants may withdraw from the intervention but continue with follow-up; withdraw from follow-up but allow data collected to date to be used; withdraw from follow-up and withdraw consent for data collected to date to be used; or be lost to contact/follow-up. Some clarification within the SAP about how each level of withdrawal will be categorised and presented is important." [1]

Example:

"The level of consent withdrawal will be tabulated and reported as a line listing containing the requisite dosing information (e.g., dose cohort assigned) and will be classified as:

- Consent to continue follow-up and data collection,
- Consent to continue data collection only,
- Complete – no further follow-up or data collection" [21]

Withdrawal/follow-up

Item 24b:

Timing of withdrawal/lost to follow-up data.

Explanation:

"Timing of withdrawals and lost to follow up is important information. This information allows you to see if there are any patterns in lost to follow up or withdrawals between the different time points and dosing schedules/sub-groups. Timing of withdrawal from follow-up or lost to follow up data can be presented in a Kaplan-Meier graph, a table or incorporated into a CONSORT flow diagram. For each follow-up time point information on the number of withdrawals and reasons for withdrawal, number included in the analysis and the number died (if applicable) should be provided." [1]

Example:

"This will be presented in tabular format, with numbers of withdrawals, discontinuations or dropouts, number of days to withdrawal, and reasons for withdrawal, drop outs or discontinuations for each Cohort, as in Table X." [21]

Withdrawal/follow-up

Item 24c:

Reasons and details of how withdrawal/lost to follow-up data will be presented.

Explanation:

"Patients can withdraw and be lost to follow up for many different reasons e.g. moved home, unable to participate any longer, withdrawn by clinician reasons etc. It is useful for the trial team to attempt to ascertain reasons for all withdrawals and loss to follow up. According to ICH E6 'Although a subject is not obliged to give his/her reason(s) for withdrawing prematurely from a trial, the investigator should make a reasonable effort to ascertain the reason(s), while fully respecting the subject's rights'. [24] Details of how this data will be presented should be included in the SAP. This information may be presented by intervention arm within a CONSORT flow diagram or in a table." [1]

Example:
"Withdrawals/loss to follow-up together with reasons will be reported by treatment schedule." [18]

Baseline patient characteristics

Item 25a:
List of baseline characteristics to be summarized.

Explanation:
Presentation of baseline characteristics is crucial for every trial as it allows the reader to see whether the characteristics are balanced across any intervention groups or consistent with the target population. Details of which baseline characteristics will be summarised in the final report should be specified along with the population for which characteristics will be presented.

If there is a randomised element to the trial, it is important to present baseline characteristics for the entire trial and by randomised treatment, and at a minimum report baseline characteristics over any factors which the randomisation has been stratified or minimised over.

For dose escalation trials, it may be preferable to present baseline characteristics by allocated dose or enrolment cohort as well as across all dose levels.

For single arm phase II trials, baseline characteristics can be presented over the entire population or by appropriate subgroup.

For trials with a suitably small sample size, it may be appropriate to report individual baseline characteristics as a line listing. If this is to occur suitable information regarding this should here be included.

Example:
"These characteristics will be presented by analysis cohort. At a minimum, this will include:

- Age at time of trauma (years),
- Total injury severity score (for the trauma cohorts only), and
- Mechanism of injury.

Further characteristics may be added as the discretion of the trial statistician, TMG, and DMC." [8]

"Baseline characteristics, including important prognostic, demographic and clinical variables will be reported overall for the main population." [18]

"Line listings will also be produced of baseline patient characteristics recorded on the Registration Form and Screening Form. Tabulated data will include: age at registration, sex, disease status at trial entry, disease history including time from first diagnosis to registration." [6]

Baseline patient characteristics

Item 25b:
Details of how baseline characteristics will be descriptively summarized.

Explanation:
"It is important to describe how baseline characteristics will be summarised and presented in the final analysis report. Formal statistical comparisons of baseline data by randomised groups are not

advocated [30, 31] but if such comparisons are planned these should be justified. It is recommended that prognostic baseline characteristics are presented for the analysis population included in the primary analysis of the primary outcome as well as for all randomised participants in order to assess whether attrition has introduced selection bias and/or upset the balance achieved at randomisation." [1]

Example:

"Baseline characteristics will be presented descriptively (without statistical hypothesis testing) on characteristics collected at the point of trial entry." [8]

Section 6: Analysis

Examples of estimands are given after all the relevant explanations have been given.

Estimand definition [b]

List and describe each primary and secondary estimands including details of:

Explanation:

The SAP should define each estimands explicitly clearly identifying primary and secondary variables. Definitions of estimands are captured in 26a-e based on ICH E9 (R1) which details the estimand framework that has been adopted by various clinical trial regulators. [32, 33]

Estimand definition [b]

Item 26a:

Treatment (including treatment combinations).

Explanation:

Details regarding the treatment of interest and, if applicable as in the instance of a randomised phase I, any alternative treatments to which comparisons will be made. In the context of dose-escalation trials where multiple doses may be under investigation, it should be made clear if the outcome will analyse patients according to their cohort, dose received, pooled across all dose levels, or some combination of the aforementioned.

Estimand definition [b]

Item 26b:

Population.

Explanation:

The trial population, defined with reference to item 20, pertinent to the outcome should be clearly stated.

Estimand definition [b]

Item 26c:

Variable of interest.

Explanation:

The endpoint to be obtained for each patient that is required to address the scientific question. If an

outcome is recorded at multiple time points, it should be made clear which of these time points are required for the specific outcome. Detailed explanations should be provided, for example for survival outcomes making it clear what the length of survival is (e.g., calculated from the time of randomisation or time of administration of intervention) and censoring information. Details given here should include specific measurements and units, especially pertinent when multiple collection methods are used. Details need to be provided on what data manipulations or derivations will be performed and how they will be carried out (e.g., change from baseline, Quality of Life (QoL) score, Time To Event (TiTE), logarithmic transformations). If the calculation of a score is more complex, but a validated algorithm is available, then providing a reference and a link to the algorithm is sufficient. Scoring, including handling of missing data, should follow guidance proposed by the instrument developers, unless there is good reason to use an alternative technique, which should be described and justified. Sufficient detail needs to be provided in order for the reader to understand how the scores or results are to be calculated for each outcome.

For dose escalation trials where dose escalation is dependent on observed rates of dose limiting toxicities (DLTs), specification of (or suitably reference to) the definition of a DLT, its reporting window and how the maximum tolerated dose (MTD) and the recommended phase II dose (RP2D) will be identified.

Estimand definition [b]

Item 26d:

Intercurrent event handling strategy.

Explanation:

Intercurrent events of interest should be defined here. Details regarding the strategy, including analysis adjustments, for dealing with intercurrent events should be specified. The five strategies for handling intercurrent events are: the treatment policy strategy; the composite strategy; the hypothetical strategy; the principal stratum strategy; and the while on treatment strategy. These strategies can be used independently or in combination, but intention of how to use should be clearly specified in advance of any analysis. [32]

Estimand definition [b]

Item 26e:

Summary measures.

Explanation:

Indication as to the population–level summary measure of the variable to which will be used. The summary measure provides a basis for a comparison between treatments or doses.

Examples

The estimand is described by the following attributes:

[26a] Treatment: Drug X infusions at days 1, 8 and 15 at a dose specific to the entry cohort as recommended by the CRM design.

[26b] Population: The evaluable population as defined in item 20. [The evaluable patient population is defined as those who meet the eligibility criteria, at a minimum have received the day 1 infusion and excludes those who have withdrawn for non-treatment related reasons.]

[26c] Variable of interest: Incidence of dose limiting toxicity (DLT) within the first 8 days of treatment. A DLT will be any adverse event (categorised as per CTCAE) which is graded as severe (grade 3) or higher and is deemed to be at least potentially related to treatment. Any patient who withdraws or dies due to treatment related reasons will be categorised as having experienced a DLT.

[26d] The following intercurrent events (IEs) of interest will be considered:
(1) Day 8 toxicity assessment not performed through patient related reasons.
(2) Day 8 toxicity assessment not performed due to site error.
(3) Day 8 toxicity assessment not being performed at the right time (performed either earlier or later than scheduled).

For IE (1), the reasons why the assessment was not performed will be investigated. Depending on the reasons for non-attendance a decision will be made regarding whether they are to be:

- Included in the analysis and assumed to have experienced a DLT;
- Included in the analysis and assumed to not have experienced a DLT; or
- Excluded from analysis and replaced with recruitment of additional patient.

For intercurrent event (2) data from subsequent visit(s) will be used to ascertain if a suspected DLT occurred during the DLT reporting window. The main estimand will use all patients who had their day 8 assessment and those who it can be definitely ascertained to have experienced a DLT within the report window (using data from subsequent visits). Any patient who did not have the day 8 assessment and who either did not experience a DLT or experienced a DLT outside of the reporting window will be excluded from the analysis.
The sensitivity estimand will then include the entire population as defined above, therefore covering all those as in the population who both did and did not have their day 8 assessment performed. For patients who missed the day 8 the following will hold: any patient who experiences an event which fulfils the criteria of a DLT at any point up until their safety visit will be assumed to experience a DLT; any patient who does not experience an event fulfilling the criteria of a DLT at any point up until their safety visit will be assumed to not experience a DLT at any point.

For IE (3) an analogous approach to the strategy defined to handle IE (2) will hold. Where it is the case that the safety assessment occurs prior to completion of the DLT reporting window, then data will also be ascertained from the first safety visit occurring after the completion of the DLT reporting window.

[26e] Summary measure: The number (count), proportion and percentage of patients experiencing a DLT per dose cohort. The estimated DLT probability for each dose from the CRM model, and the subsequent recommended dose.

The primary estimand is described by the following attributes:

[26a] Treatment: 7 infusions of drug X at dose Y mg/kg approximately 7 days apart starting on day 1.

[26b] Population: The modified intention to treat (mITT) population as defined in item 20. [The mITT population contains all patients who have received at least one infusion at the confirmed dose of drug X.]

[26c] Variable of interest: Serum alkaline phosphatase (ALP) at visit 3 (pre-infusion) and at follow up visit 10 (day 99) as evaluated at central laboratory.

The primary estimand is the response at day 99 in serum ALP, requiring a reduction of 25% or more from baseline. Baseline ALP level will be measured at pre-infusion on the first treatment visit (overall trial visit 3), and again at follow-up visit 10 day 99. The response will be calculated using the formulae:

$$\%\ Change = \frac{ALP_{Day99} - ALP_{Visit\ 3\ (pre-infusion)}}{ALP_{Visit\ 3\ (pre-infusion)}} \times 100$$

Using the above formulae, a negative value indicates a reduction, whereas a positive value indicates an increase.
The clinically meaningful reduction required corresponds to a value of -25% or less (≤ -25%). The proportion of patients with a clinically meaningful reduction will be calculated as

$$\%\ of\ patients\ with\ clinically\ meaningful\ reduction = \frac{Patients\ with\ clinically\ meaningful\ reduction}{All\ patients\ in\ the\ mITT\ population}$$

Where patients have their follow up visit 10 ALP sample missing, they shall be treated as a non-responder and included in the denominator of the above equation. The number of non-responders for the primary outcome will be reported.

[26d] Intercurrent event: The key intercurrent events pertains to blood samples not being analysed or returned from the central laboratory (e.g., due to samples haemolysing or being lost in transit). In order to mitigate against this further samples will be analysed locally. It is our intention to use the principal stratum strategy, and thus only analyse patients who have centrally analysed samples in the primary estimand.

[26e] Summary measure: The number and proportion of patients with a clinically meaningful reduction will be reported.

Analysis methods

Item 27a:

What estimator and analysis method will be used and how the results will be presented.

Explanation:

Conclusions can be affected substantially by the analysis method(s) used, therefore, it is extremely important to pre-specify the analysis method(s) so there is no possibility of the method being chosen because it gives the most positive results. For each outcome, the SAP should specify what analysis method(s) will be used for statistical comparisons. The population and summary measure used should be consistent with that specified in the definition of the estimand, in items 26a-e. If more than one method is to be used to analyse the primary outcome, e.g., adjusted and unadjusted for covariates, then the primary analysis method should be identified.

Where line listings are to be used, it may be prudent to here include which information will be reported.

For dose escalation trials, the criteria for deciding to escalate doses and how the final dose will be picked (e.g., that with DLT probability closest to but not exceed 33%) should be described.

For all model-based and model-assisted early phase trials it is useful to include the formulae (or sufficient reference to), and the mathematical specification of the model required for the analysis. If these formulae have been specified in earlier sections, such as in item 9 or in a supporting document, reference to this is sufficient. Moreover, if transformations are to be applied, then these should be specified along with the rationale for the transformation and the resulting interpretation.

To ensure that critical decisions and conclusions drawn from the trial where the analysis method is novel or non-conventional, it is recommended that the code required to produce the analysis and, where appropriate, inform dose escalation decisions is made available. While the main body of the SAP is not the appropriate place for this, it is suggested that the code is appended, see point 35. Making the code available allows the critical decisions of the trials to be replicated and reproduced.

Example:

"The estimator in order to determine the optimal dose is the EffTox design (as described in item 9c). The optimal dose will be reported with its associated probability of DLT and response."

"We will use a 3-parameter logistic regression model to model the relationship between dose and efficacy. Patients are assessed for the efficacy endpoint in week 13. Patients who have not reached this time point yet will not provide any information to this model. Patients who have reached the time point for this endpoint and did not have the evaluation, or who withdrew or died prior to evaluation will be treated as non-responders.
Let z_j denote 1 if the patient responded and 0 otherwise. Then we assume

$$z_j|x_j \sim Bernoulli(\emptyset_{x_j})$$

$$logit(\emptyset_{x_j}) \sim \gamma_1 + \gamma_2 \log\left(x_j/d^*\right) + \gamma_3\left[\log\left(x_j/d^*\right)\right]^2$$

$$\pi(\gamma_1, \gamma_2, \gamma_3) \sim MVN\left(\begin{pmatrix} -0.5 \\ 0.5 \\ 0 \end{pmatrix}, \begin{pmatrix} 7 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & 4 \end{pmatrix}\right)$$

We can then calculate the posterior probability of efficacy for each treatment schedule." [18]

"The estimand is estimated by the number and proportion of patients with a clinically meaningful ALP reduction as described in item 26e."

Analysis methods

Item 27b:

Any adjustments for covariates.

Explanation:

For each estimator which has an underlying statistical model, the SAP should specify whether adjustment will be made, and if so, the covariates to be used (including the categories if applicable), and how these will be included in the model (e.g., as fixed effects, or random effects). For the primary endpoint, it must be clear whether the adjusted or unadjusted analysis is the primary analysis as failing to pre-specify can lead to bias.

Example:

"Baseline covariates will be adjusted for in the modelling as necessitated by clinical indication and in order to aid model convergence/diagnostics."

"No adjustments for covariates will be made."

Analysis methods

Item 27c:

Methods used to check assumptions of the underlying statistical methods and goodness of fit for the model.

Explanation:

For each estimator which has necessary post estimation check, there may be a number of assumptions which need to hold for the analysis to be valid and to ensure that conclusions, and where appropriate dose escalation decisions, drawn are correct. Checks to assess the underlying assumptions should be pre-specified.

Example:

"The first method of checking model adequacy will be the presence of divergent transitions. Presence of any divergent transitions will indicate that the proposed model does fit the observed data satisfactorily, and that alternative models need to be considered.
First, alternative specifications for any fixed-effects will be considered. Analytical functions of time-varying covariates will be considered (e.g., Time^2, or $\sqrt{Time}$) to address the potential of non-linear progression. Secondly, and where appropriate, alternative specifications for the random-effects will be considered. It is anticipated that the terms used in the random effects structure will be a subset of those used in the fixed effects structure. In all situations, a saturated model is likely to provide a good fit. However, we will prefer a more efficient model with fewer parameters, if possible. In all cases, the final functional form of the models used will be presented.
While all models will be run on multiple chains, and a warm-up sample discarded with the aim of minimising the possibility of non-convergence, non-convergence is possible. Model convergence will be assessed through visual inspection of history, density, and autocorrelation plots. Model convergence statistics of and the effective sample size will also be monitored. As with all convergence plots, such methodology is only appropriate for detecting non-convergence, should any of the aforementioned convergence plots or statistics suggest evidence of non-convergence, sensitivity to warm-up and sample, inclusion of different baseline covariates, and alternative model specifications will be considered." [34]

"Given that only descriptive statistics are to be presented, there is no appropriate method for checking assumptions. The type of diagnostic statistics (either means and SDs or medians and IQRs) will be chosen based on the distribution of observed data."

Analysis methods

Item 27d:

Details of alternative methods to be used if distributional assumptions do not hold.

Explanation:

Since randomisation and blinding can be rare in early phase trials, a blinded review of distributional assumptions may not be relevant or possible. Therefore, it is important to pre-specify alternative methods and models which are to be used if the underlying assumptions do not hold. Akin to the main estimator and where possible, the formulae and mathematical specification of these alternative models should be given.

The approach taken should be considered carefully as bias may be introduced either by choosing the method of analysis based on the results of tests of assumptions [35, 36] or from performing hypothesis tests in which the underlying assumptions are not upheld. Three possible approaches may be considered: i) pre-specify alternative analyses and how the statistician will choose between them in the SAP so that the process is transparent; ii) select a method of analysis that is robust to assumptions; or iii) state the method of analysis to be used in the SAP and specify that a sensitivity analysis will be performed using an alternative set of assumptions and the results compared.

Example:

"The relationship of the primary endpoint with dose will be explored by fitting a number of candidate models. The list of candidate models were fitted to the primary endpoint divided by 100 with the targets defined as 0.10 and 0.20, respectively. The candidate models include the linear, the quadratic, the cubic, the logistic and the Emax (with 3 and 4 parameters). The mathematical formula of the models are given by:

- Linear: $y = a + b\ (dose)$
- Quadratic: $y = a + b\ (dose) + c\ (dose^2)$
- Emax: $y = \{ a + b\ (dose) \} / \{ 1 + dose/ c\}$
- Cubic: $y = a + b\ (dose) + c\ (dose^2) + d\ (dose^3)$
- Logistic: $y = a + \{ b / 1 + \exp ((c - dose)/ d) \}$
- Emax4: $y = \{ a + b\ dose^{\exp(d)} \} / \{ 1 + dose^{\exp(d)} / c\}$

The estimated equation of each convergent model will also be plotted in the scatter plot of the primary endpoint against dose. For each model, its estimated coefficients (a to d where applicable) with their standard error will be reported. The Akaike Information Criterion (AIC) and the deviance of each model will be reported as measures of adequacies of fit. The residual error of each regression will also be reported. The residual values of each model against its predicted values will be plotted, as well as a quantile-quantile plot of its residuals. The proposed target doses of each model with their standard errors and with a 95% confidence interval will be reported." [7]

"*A priori* it is not thought that any alternative model specifications will be warranted. If this ends up not to be true, the alterations to this will be detailed in all generated reports and marked as a deviation from the SAP".

"No modelling is here to be performed and so specification of alternative models is not appropriate".

Analysis methods

Item 27e:

Any planned sensitivity analyses for each outcome where applicable.

Explanation:

For each outcome, where applicable and in line with the definition of the estimand, the SAP should specify whether any sensitivity analyses will be conducted. The definition and description of any planned sensitivity analyses should include the same level of detail as in the descriptions of the primary and secondary estimators. Any parts of the estimand which will change when conducting sensitivity analysis (e.g., a change in analysis population) should be clearly defined and explained. Moreover, while it is unlikely in the context of early phase clinical trials that the presence of a high amount of missing data would trigger the need for sensitivity analyses, if such a minimum percentage does exist, this should be clearly stated.

Example:

"There are two planned sensitivity estimands to the primary estimand planned. For the first, where patients do not have a centrally analysed ALP value, the locally analysed ALP value will be imputed and the primary estimand repeated. The second sensitivity analyses will repeat the primary estimand this time using only locally analysed ALP for all patients. This sensitivity analyses is intended to only be exploratory and so no significance testing will be performed on sensitivity results."

"Sensitivity analyses will also be performed in the per-protocol population, which is defined as those patients who completed the treatment as originally allocated with no dose modification or missing doses (i.e. patients that have received all 7 infusions as scheduled in the protocol at the MED dose). For sensitivity analyses, only the primary outcome measure (with central processing) will be assessed." [3]

"Sensitivity analyses will be carried out during the trial for dose-decision meetings and also for final analysis for estimating the optimal dose. In addition to repeating the analysis using the sensitivity population defined in Section 4.2, we will also repeat the analysis using different weight functions. Therefore the 2 sensitivity analyses are:

1. Sensitivity population and analysis using weights according to length of follow-up only, i.e. not taking into account how much dose has been received
2. Main population but with the most toxic scenario, i.e. we assume that all patients currently in follow-up within the DLT Window of 13 weeks have a DLT" [18]

"For all the Bayesian analysis listed above, where prior distributions are specified in advance, sensitivity to prior will be assessed." [34]

"There are no planned sensitivity analyses for this study." [21]

Analysis methods

Item 27f:

Any planned subgroup analyses for each outcome including how subgroups are defined.

Explanation:

All pre-planned subgroup analyses should be clearly specified. It may be appropriate to define the subgroup analysis using the estimand framework, including the same considerations such as how the patient populations will be defined and patients assigned subgroup categories, and how the results will be presented. Performing a large number of subgroup analyses is often infeasible in early phase trials due to the limited sample size and should generally be avoided. However, there may be times when it is appropriate to do so (for example when the aim is to demonstrate consistency across subgroups).

Example:

"Due to the lack of statistical power for subgroup analyses in this early phase II trial, results provided will be exploratory only. Therefore, results should not be over interpreted and instead used as a guide for further subgroup analyses in a larger phase III setting. Subgroups to be studied include:

- Results by mutation type
- Primary disease
- Oestrogen receptor data

- Sex
- Health Economics" [37]

"Exploratory subgroup analyses may be performed based on stage of liver disease and/or prior treatments. For exploratory subgroup analyses, only the primary estimand (with central processing) will be assessed and no hypothesis testing performed." [3]

"No subgroup analyses are planned." [2]

Missing data

Item 28:

Reporting and assumptions/statistical methods to handle missing data (e.g., multiple imputation).

Explanation:

While the majority of trials will have some missing data, [38] thus potentially introducing bias dependent on the pattern of 'missingness', [39] using formal methods such as multiple imputation to handle missing data is generally not advocated in early phase clinical trials due to the restricted sample size.

Regardless, it is important that the SAP states how missing data will be handled and reported including details of any statistical methods and their assumptions, which will be used to handle missing data. The definition of what data is considered to be missing and the methodology used to deal with any missing data will be directly impacted by the definition of the estimand, as given in items 26.

If there are plans to impute missing outcome data, then a list of variables and details regarding the imputation process should be made apparent, either through explanation here or through suitable reference to another supporting document where more details can be sought. Since conclusions drawn from any imputation depend on the statistical method used, it is crucial to pre-specify which methods will be used under which circumstances, and which will be considered the primary analysis. This again promotes transparency in the trial and minimises any ambiguity in the methods.

Example:

"The incident of missing data will be reported and if it rises more than 10% then sensitivity analyses will be carried out as appropriate. Summary tables will present the population size either in the title or in the column headers, and thus the number of missing values for any particular variable/visit will be documented." [7]

"No data imputation is planned." [18]

Additional analyses

Item 29:

Details of any additional statistical analyses required, e.g., complier-average causal effect analysis.

Explanation:

"Any additional analyses to be conducted should be specified with reasons these are required, a description of the additional analysis and how it will be conducted. This may include pre-specified

exploratory analyses that are hypothesis generating or confirmatory of issues identified in other trials." [1]

Example:
"No imputation or additional analyses are planned *a priori.*" [9]

"Additional analyses will be performed which combines translational data with clinical outcomes."

Harms

Item 30:
Sufficient detail on summarizing safety data outside of that used for dose escalation (e.g., non-DLT safety data), e.g., information on severity, expectedness, and causality; details of how adverse events are coded or categorized; how adverse event data will be analysed, i.e., by grade, incidence case analysis, intervention emergent analysis.

Explanation:
Where information on DLTs is collected, incidence and details of DLTs will typically be reported alongside the relevant outcome. However, consideration of the full safety profile is key for every clinical trial. It is important that safety data is reviewed and details are provided in the SAP on how the remaining safety data will be summarised during interim and final analyses, including the analysis population to be used. Information may be provided on the severity, causality and expectedness of the adverse event, events resulting in dose reductions, information on how the adverse events will be coded or categorised and by whom. The method of summarising the adverse event data should be described ensuring it is clear whether the descriptive summary will use number of events or number of patients and any analyses to be conducted (e.g., will the adverse events be compared descriptively or will formal statistical testing be undertaken).

Example:
"All safety analysis will be conducted on the safety analysis set.
In order to assess toxicity throughout the trial, the following will be presented at each DMC meeting and in the primary analysis report.

- The number of deaths in the trial will be reported by cohort and treatment arm with cause of death reported.
- The number of serious adverse events (SAEs) will be presented by cohort and treatment arm. A summary for each SAE categorisation code (e.g., SAR, SUSAR, unrelated SAE) will be presented.
- The number of grade 3 or greater adverse events reported by cohort and treatment arm (for the randomised component).
- The number and proportion of patients experiencing a grade 3 or greater adverse event.
- The number and proportion of patients experiencing any adverse event." [22]

"Details of all AEs will be documented and reported from the date of commencement of protocol defined treatment until 28 days after the administration of the last treatment. All AEs will be followed up until resolution or until last trial visit (whichever occurs soonest). Any AEs ongoing at the patient's last trial visit will be marked as unresolved.
In addition to safety outcomes as detailed below, the following will be reported, stratified per-protocol and, where appropriate, according to trial stage.

- Toxicities will be tabulated by CTCAE v5.0 classification, grade and number (and percentage) of patients affected,
- A line listings given of grade 3, 4 or 5 adverse events deemed at least possibly related to treatment,
- Duration of adverse events will be summarised,
- SAEs will be reported as frequency and number of patients experiencing them, together with outcome (e.g., death, resolved etc.), and
- Line listings of all SAEs and DLTs.
- Incidence and summary characteristics of adverse events of particular interest shall be reported. Adverse events that are of particular interest are those which are perceived to be related to tolerability of the gel, e.g., redness or itchiness of the wound." [34]

Statistical software

Item 31:

Details of statistical packages to be used to carry out design, simulation and analyses

Explanation:

"Details of the statistical packages to be used to conduct the statistical analyses may be provided in the SAP. While version numbers of software may change during the lifetime of the trial and so should not be specified in the SAP they should be included within final reports." [1]

Example:

"Statistical analyses will be carried out using relevant statistical software; SAS, Stata or R." [11]

References

Item 32a:

References to be provided for nonstandard statistical methods.

Explanation:

"References should be provided in a SAP for any non-standard statistical methods that will be used. If there is any doubt on whether a method is non-standard then it is better to include a reference." [1]

References

Item 32b:

Reference to Data Management Plan.

Explanation:

"Reference should be made to the Data Management Plan (DMP) with the version number that was used when writing the SAP. This is important as both documents should be linked with information in the DMP that is also important for the final analysis report. If there is no DMP, then the location of this information (e.g., data handling and cleaning) should be provided." [1]

References

Item 32c:

Reference to the Trial Master File and Statistical Master File.

Explanation:

"The Statistical Master File is part of the Trial Master File but is often held separately with restricted access. The Statistical Master File may hold details of the randomisation process or specific protocol deviations that the statistician needs to refer to when executing the statistical analysis plan. If a Statistical Master File is held separately to the Trial Master File, then both should be referenced." [1]

References

Item 32d:

Reference to other standard operating procedures or documents to be adhered to.

Explanation:

"Reference should be made to any other Standard Operating Procedures (SOPs) or documents that are adhered to and followed when writing the SAP." [1]

Section 7: Suggested SAP Appendices

Simulation Report

Item 33:

Operating characteristics of the trial design to assess the probability of trial success under different plausible scenarios.

Explanation:

The estimand and scenarios assessed through simulations will be analogous to the estimand used in the main trial, appraising all the underlying assumptions and limitations of the model.

Where model-based designs are used, assessment of the design's operating characteristics is needed to ensure the trial will yield a (successful) result and provide sufficient overdose control. [40] Proper documentation of simulation studies is favoured by regulators.

The appropriate simulations to assess utility of the trial design should, at a minimum, test the scenarios where each dose level is the maximum tolerated dose (MTD) and where all doses are ineffective or dangerous (e.g. too toxic or fail to achieve the desired response). With regards to simulation output for dose-escalation trials, for each scenario it is preferable to report:

- Prior distributions or skeletons (as appropriate),
- The true DLT rate,
- The probability of each dose being selected as the MTD (where applicable),
- The percentage or mean number of patients being treated at each dose level,
- The probability or mean number of patients being treated above the true and estimated (where different) MTD,
- The probability the trial stops early due to excess toxicity (e.g., when the lowest dose is too toxic).

With regards to simulation output for efficacy estimands, for each scenario it is preferable to report:

- Prior distributions (as appropriate),
- The true efficacy or response rate (as appropriate),
- For trials with a formal interim, probability of stopping for either efficacy or futility (depending on trial design), where the underlying truth would both agree with and counter indicate this,
- Probability of yielding a successful result at the end of the trial.

Sufficient information should be included to allow for replication by someone without prior knowledge of the trial. The methodology and rules underpinning the design (e.g., doses under investigation and their order, target dose, model type and parameters, including where appropriate, model weights, and how to select a dose) should be the same as for the main model to be used for analysis (as specified in the main SAP). In addition to output and results, at a minimum the following would be recommended:

- The estimated duration of the trial,
- The number of patients to be enrolled per cohort and in total,
- If a flexible cohort size is to be used, how this cohort size was sampled,
- How many simulations are run,
- If the trial design has a TiTE component, the time between patients are enrolled,
- If a seed was used, this should be included.

Further guidance, particularly for CRM trials is available. [41]

Dose transition pathways

Item 34:

For dose-escalation trials, indication of the dose transition pathways (either using tables or trees/graphs) under different DLT scenarios.

Explanation:

For any dose escalation component of early phase trials, dose transition pathways (DTPs) are a useful tool used to assist decision-making, particularly useful in the instance of novel methodology or complex dose escalation/overdose control rules. DTPs facilitate transparency of dose escalation decisions and can be a useful tool to facilitate the work of the relevant safety monitoring board. [42]

This section is not applicable for single arm phase II trials with efficacy outcomes.

Example

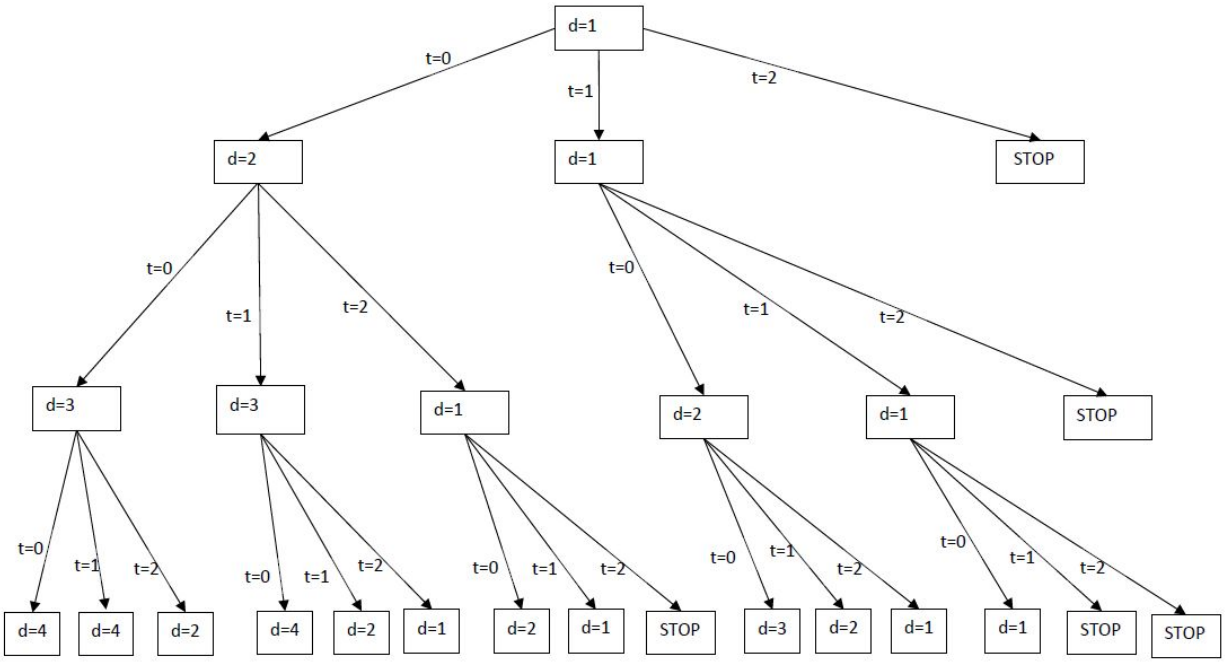


Figure A3b: Dose Transition Pathways

[18]

Table A3h: Dose transition pathways for the first four cohorts of three patients. This chart should not be used if the size of any cohort is different to three patients.

Pathway	Cohort 1		Cohort 2		Cohort 3		Cohort 4
	Dose	DLT	Dose	DLT	Dose	DLT	Dose
1	2	0	3	0	4	0	5
2	2	0	3	0	4	1	4
3	2	0	3	0	4	2	4
4	2	0	3	0	4	3	3
5	2	0	3	1	3	0	4
6	2	0	3	1	3	1	3
7	2	0	3	1	3	2	2
8	2	0	3	1	3	3	1
9	2	0	3	2	2	0	3
10	2	0	3	2	2	1	2
11	2	0	3	2	2	2	1
12	2	0	3	2	2	3	1
13	2	0	3	3	1	0	2
14	2	0	3	3	1	1	1
15	2	0	3	3	1	2	1
16	2	0	3	3	1	3	0
17	2	1	2	0	3	0	4
18	2	1	2	0	3	1	3
19	2	1	2	0	3	2	2
20	2	1	2	0	3	3	1
21	2	1	2	1	2	0	3
22	2	1	2	1	2	1	1
23	2	1	2	1	2	2	1
24	2	1	2	1	2	3	1
25	2	1	2	2	1	0	1

[6]

Code

Item 35:

Full model specification and programming code used for evaluation of dose-escalation decisions.

Explanation:

Optional section, encouraged for novel model-assisted and model-based phase I designs. In these instances, the full model specification and programming code should be made available in the SAP (or suitably referenced document). If model code is to be included, appropriate annotation of the model code should be incorporated. Where established methodology is to be used, for which there are publicly available specialist software available, appropriately referenced indication of functions and packages to be used (including an example of such functions) is sufficient. This allows dose escalations decisions to be replicated and promotes reproducibility.

Reports Template

Item 36:

Optional section detailing exemplar tables, graphs and report templates.

Explanation:

While not necessary, a template may be appended to the SAP to aid in producing reports to be used

over the course of the trial (for example for interim safety meetings; dose escalation decision reviews; or data monitoring committee meetings). These reports may detail the intended layout, content, tables, and graphs to be produced. This template may be stored separately to the SAP, in which case suitable reference is sufficient. Suggested sections in these report templates include;

- CONSORT diagram,
- Recruitment details,
- Baseline characteristics,
- Patient disposition,
- Treatment (received, discontinuation, compliance),
- Adverse Events (AEs), to include sections on DLTs, Serious Adverse Events (SAEs), non-serious AEs,
- Dose-escalation content e.g., Modelling output, recommendation for next cohort (where appropriate),
- Efficacy estimands (where appropriate),
- PK estimands (where appropriate),
- Sensitivity analysis results (different contributions to TiTE component),
- Other estimands (where appropriate).

Notes

a. *This item was called 'Confidence intervals and P values' in the Gamble et al. paper. It has been changed to 'Indications of uncertainty' to reflect that many early phase trials designs are underpinned by Bayesian methodology.*
b. *This item was called 'Outcome definitions' in the Gamble et al. paper. It has been changed to 'Estimand definitions' following the wider adoption of ICH E9 (R1; Addendum on estimands).*

Abbreviations List

- AE: Adverse Events
- AIC: Akaike information criterion
- ALP: Alkaline Phosphatase
- AUC: Area Under the Curve
- BID: Bis In Die (twice a day)
- CI: Confidence Interval
- CONSORT: CONsolidated Standards Of Reporting Trial
- CrI: Credible Interval
- CRM: Continual Reassessment Method
- CTCAE: Common Terminology Criteria for Adverse Events
- DDC: Dose Determination Committee
- DLT: Dose Limiting Toxicity
- DMC: Data Monitoring Committee
- DMP: Data Management Plan
- DTP: Dose Transition Pathways
- ESS: Effective Sample Size
- ICH: International Council for Harmonisation
- IE: Intercurrent Event
- IMP: Investigational Medicinal Product

- IQR: Interquartile Range
- ITT: Intention To Treat
- IV: Intravenous
- LPLV: Last Patient Last Visit
- MED: Minimum Effective Dose
- mITT: Modified Intention To Treat
- MRD: Minimum Residual Disease
- MTD: Maximum Tolerated Dose
- p_0: The largest unacceptable response rate
- p_1: The smallest acceptable response rate
- PD: Pharmacodynamics
- PK: Pharmacokinetics
- PO: Partial Ordering
- QoL: Quality of Life
- RP2D: Recommended Phase II Dose
- SAE: Serious Adverse Event
- SAP: Statistical Analysis Plan
- SAR: Serious Adverse Reaction
- SD: Standard Deviation
- SOP: Standard Operating Procedures
- SUSAR: Suspected Unexpected Serious Adverse Reaction
- TiTE: Time To Event
- TMG: Trial Management Group
- TSC: Trial Steering Committee

References

[1] Lifted verbatim from: Gamble C, Krishan A, Stocken D, et al. Guidelines for the Content of Statistical Analysis Plans in Clinical Trials. JAMA. 2017;318(23):2337–2343. doi:10.1001/jama.2017.18556

[2] Elhussein L, Moschandreas J, Khan I, Vyas P.CAMELLIA. Full title: A Phase I dose escalation trial of the Humanized Anti-CD47 Monoclonal Antibody Hu5F9-G4 in Haematological Malignancies. Statistical Analysis Plan. [Unpublished SAP]

[3] Homer V, Slade D, Kirkham A, Fox R, Newsome P, Hirschfield G. A single arm, two-stage, multi-centre, phase II clinical trial investigating the safety and activity of the use of BTT1023, a human monoclonal antibody targeting vascular adhesion protein (VAP-1), in the treatment of patients with primary sclerosing cholangitis (PSC). Statistical Analysis Plan. [Unpublished SAP]

[4] De Angelis C, Drazen JM, Frizelle FA, et al. Clinical trial registration: a statement from the International Committee of Medical Journal Editors. New England Journal of Medicine 2004;351(12) :1250-51.

[5] Tovey H, Hall E, Yap T. ComPAKT: A Phase I multi-centre trial of the Combination of olaparib (PARP inhibitor) and AZD5363 (AKT inhibitor) in patients with advanced solid tumours (CCR4058) Statistical Analysis Plan. [Unpublished SAP]

[6] Slade D, Yap C, Collins G. RomICar. Phase I/II study to determine the maximum tolerated dose and activity of the combination of romidepsin and carlzomib in relapsed or refractory peripheral T-cell lymphoma. Statistical Analysis Plan. [Unpublished SAP]

[7] Evangelou M, Bond S, Waldron-Lynch F. DILT1D. Adaptive study of IL-2 dose on regulatory T cells in type 1 diabetes. Statistical Analysis Plan [Unpublished SAP]
[8] Homer V, Slade D, Brock K, Foster M. ADaPT. A prospective, phase II, single centre, cross-sectional, randomised study investigating Dehydroepiandrosterone and Pharmacokinetics in Trauma. Statistical Analysis Plan. [Unpublished SAP]
[9] Munoz Vicente S, Brock K, Hillmen P. CLARITY: Assessment of VenetoCLAx (ABT-199) in combination with IbRutInib in relapsed/refracTory Chronic LymphocYtic Leukaemia. Statistical Analysis Plan. [Unpublished SAP]
[10] Patel A, Gaunt P, Mehanna H, Kong A. ADePT DDR. Accelerating the Development and implementation of Personalised Treatments of DNA Damage Response agents and radiotherapy +/- immunotherapy for head and neck squamous small cell cancer. Statistical Analysis Plan. [Unpublished SAP]
[11] Patel A, Billingham L, Patel P. AdUp. A Phase I Clinical Trial of a replication defective type 5 adenovirus vector expressing nitroreductase and GMCSF (AdNRGM) given via trans-perineal, template-guided, intra-prostatic injection, followed by intravenous CB1954, in patients with locally relapsed Prostate Cancer. Statistical Analysis Plan. [Unpublished SAP]
[12] Homer V, Slade D, Brock K, Srinivasan V. ALS: A Phase II pilot safety and tolerability study of ILB in patients with Motor Neurone Disease (MND) / Amyotrophic Lateral Sclerosis (ALS). Statistical Analysis Plan. [Unpublished SAP]
[13] Thall PF and Cook, JD, Dose-Finding Based on Efficacy–Toxicity Trade-Offs. Biometrics, 2004. 60: p. 684–693
[14] Cook, JD. Efficacy-Toxicity trade-offs based on L^p norms. Technical Report UTMDABTR-003-06
[15] Thall, PF et al. Effective sample size for computing prior hyperparameters in Bayesian phase I–II dose-finding. Clinical Trials. 2014. 11(6). p.657-666
[16] Brock K, Yap C, Slade D, Copand M. MATCHPOINT: MAnagement of Transformed CHronic myeloid leukaemia: POnatinib and INTensive chemotherapy: a dose-finding study. Statistical Analysis Plan [Unpublished SAP]
[17] Kirkham A, Yap C, Mehanna H. WISTERIA. A Phase I trial of WEE1 inhibition with Chemotherapy and Radiotherapy as adjuvant treatment, and a Window of Opportunity trial with Cisplatin in Patients with Head and Neck Cancer. Statistical Analysis Plan. [Unpublished SAP]
[18] Holmes J, Ooms A, Hawkins M. Cedar. Full title: A Phase 1 study of the safety, tolerability and biological effects of intravenous EnAdenotucirev, a novel oncolytic virus, in combination with chemoradiotherapy in locally advanced rectal cancer. Statistical Analysis Plan. [Unpublished SAP]
[19] Ali A, Billingham L, Menne T. SeluDex. International phase I/II expansion trial of the MEK inhibitor selumetinib in combination with dexamethasone for the treatment of relapsed/refractory RAS-pathway mutated paediatric and adult Acute Lymphoblastic Leukaemia. Statistical Analysis Plan. [Unpublished SAP]
[20] Lloyd S, Li L, Zhou, Sampson M. A Phase 1, Open-Label, Randomised, Repeat Dose, Parallel Group Study to Evaluate the Pharmacokinetics, Safety and Tolerability of Ferric Maltol at Three Dosage Levels in Paediatric Subjects Aged 10-17 Years of Age with Iron Deficiency (with or without Anaemia). May 2018. Accessed online at https://clinicaltrials.gov/ProvidedDocs/51/NCT03181451/SAP_001.pdf, 7th July 2020
[21] Wilhelm-Benartzi C, Casbard A, Chester J. TOTEM. Trial of Temsirolimus for Advanced Cancers. A Phase I/II single-arm trial to evaluate the combination of cisplatin and gemcitabine with the mTOR inhibitor temsirolimus for treatment of advanced cancers,

including first-line treatment of patients with advanced transitional cell carcinoma of the urothelium. Statistical Analysis Plan [Unpublished SAP]

[22] Boucher R, Jackson A, Fox S, Schuh A. STELLAR. A phase II, randomiSed study of CHOP-R in combination with acalabruTinib comparEd to CHOP-R in patients with newLy diagnosed Richter's Syndrome (RS) and a pLAtfoRm for initial investigations into activity of novel treatments in relapsed/refractory and newly diagnosed RS. Statistical Analysis Plan. [Unpublished SAP]

[23] Lifted verbatim from: Cole M, Yap C, Buckley C, Ng WF, McInnes I, Filer A, Siebert S, Pratt A, Isaacs JD, Stocken DD. TRAFIC: statistical design and analysis plan for a pragmatic early phase 1/2 Bayesian adaptive dose escalation trial in rheumatoid arthritis. Trials. 2021 Jul 6;22(1):433. doi: 10.1186/s13063-021-05384-5. PMID: 34229728; PMCID: PMC8259060.

[24] F. Hoffmann-La Roche Ltd. Study BP42698: A DOUBLE-BLIND, RANDOMIZED, PARALLEL-GROUP, PHASE 2 STUDY TO INVESTIGATE THE EFFECT OF RO7049665 ON THE TIME TO RELAPSE FOLLOWING STEROID TAPERING IN PATIENTS WITH AUTOIMMUNE HEPATITIS. [Unpublished study protocol]

[25] European Medicines Agency, "Guideline on multiplicity issues in clinical trials," Guidance, 2016.

[26] Medicines for Healthcare products Regulatory Agency. Good Clinical Practice Guide. United Kingdom: The Stationary Office, 2012.

[27] Brown S, Pawlyn C, Tillotson AL, Sherratt D, Flanagan L, Low E, Morgan GJ, Williams C, Kaiser M, Davies FE, Jenner MW; Myeloma UK Early Phase Clinical Trial Network. Bortezomib, Vorinostat, and Dexamethasone Combination Therapy in Relapsed Myeloma: Results of the Phase 2 MUK four Trial. Clin Lymphoma Myeloma Leuk. 2021 Mar;21(3):154-161.e3. doi: 10.1016/j.clml.2020.11.019. Epub 2020 Dec 3. PMID: 33478922.

[28] Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMC Medicine 2010;8(1):18.

[29] International Conference on Harmonisation. Topic E6 (R1) Guideline for Good Clinical Practice (CPMP/ICH/135/95) European Medicines Agency, 1996

[30] Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. Journal of clinical epidemiology 2010;63(8):e1-e37.

[31] Senn S. Testing for baseline balance in clinical trials. Statistics in medicine 1994;13(17):1715-26.

[32] International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use, "Addendum on Estimands and Sensitivity Analysis in Clinical Trials To the Guideline on Statistical Principles for Clinical Trials E9(R1)," Fed. Regist., 2019.

[33] European Medicines Agency, "ICH E9 - Statistical Principles for Clinical Trials," ICH Guidel., vol. NA;, no. 1532, p. 3, 2006.

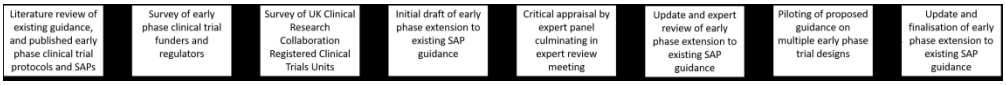
[34] Homer V, Slade D, Moiemen N. DeScar. A first-in-human clinical trial of a bioactive dressing designed to reduce scarring of skin burns. Statistical Analysis Plan. [Unpublished SAP]

[35] J. J. Shuster, "Diagnostics for assumptions in moderate to large simple clinical trials: Do they really help?," Stat. Med., 2005.

[36] H. Campbell and C. B. Dean, "The consequences of proportional hazards based model selection," Stat. Med., 2014.

[37] Boucher R, Houlton A, Sirovica M, Bishop R, Harrison C. TAMARIN. Effects of TAMoxifen on the Mutant Allele Burden and Disease Course in Patients with MyeloprolifeRative Neoplasms. Statistical Analysis Plan. [Unpublished SAP]

[38] A. M. Wood, I. R. White, and S. G. Thompson, “Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals,” Clinical Trials. 2004.
[39] D. B. Rubin, “Inference and missing data,” Biometrika, 1976.
[40] T. P. Morris, I. R. White, and M. J. Crowther, “Using simulation studies to evaluate statistical methods,” Stat. Med., 2019.
[41] G. M. Wheeler et al., “How to design a dose-finding study using the continual reassessment method,” BMC Med. Res. Methodol., 2019.
[42] C. Yap, L. J. Billingham, Y. K. Cheung, C. Craddock, and J. O’Quigley, “Dose transition pathways: The missing link between complex dose-finding designs and simple decision-making,” Clin. Cancer Res., 2017.



Caption: Figure S1. Schema detailing overview of early phase SAP guidance extension process

1474x112mm (59 x 59 DPI)

CTU Name: _____

Details of statistics representative who completed this survey:

Name: _____

Email: _____

We ask the above details for the purposes of keeping a record of responders and to ascertain aggregate data regarding survey completion. If you do not wish to be acknowledged for your participation, please tick here: ☐

Please return the completed surveys to V.S.Homer@bham.ac.uk.

The following questions are about Statistical Analysis Plans (SAPs) and current practice within your CTU. We would also be grateful if your responses could be made on behalf of your CTU, so it may help to discuss the survey with your colleagues before returning it.

For the purpose of this project and survey, we are defining early phase trials as trials which aim: to determine safe doses and dosing schedules for a treatment/intervention (phase I), or whether or not there is any signal of efficacy for that intervention (phase II or I/II).
Our definition therefore includes single-arm or randomised phase I trials and single-arm phase II trials such as:

- Rule-based phase I trials (such as the 3+3 design),
- Model-based phase I trials (such as the continual reassessment method),
- Model-assisted phase I trials (such as modified toxicity probability interval (mTPI) design),
- Randomised dose finding phase I trials (such as those which randomise between placebo and a dose of the experimental treatment, or those which randomise to attain the optimal doses or dose schedules once safety has been assured),
- Single arm phase II trials.

Does your trials unit run early phase clinical trials?

Yes ☐ No ☐

If you answered no to the above question, we would appreciate it if you could return the survey with only this question answered using the details provided at the start. We would like to thank you for your time in reviewing our request and shall no longer contact you in relation to this project.

If your unit does run early phase clinical trials, we would greatly appreciate it if you could complete the remainder of this survey.

1. Regardless of trial phase, does your trials unit have a SAP template, or a specific set of instructions, that you use when authoring SAPs?

Yes ☐ No ☐

1a. If no, reason why? Please tick one:

Template is not required ☐

Please specify why _____

Template under development ☐

Need for template recognised but development has not started ☐

Other (please specify) _____

2. When authoring SAPs for early phase clinical trials, as defined at the start of this document, which of the following is most appropriate:

a. we have a specific SAP template (or set of instructions) for early phase clinical trials ☐

b. we have a SAP template (or set of instructions) that is applicable to all phases of trial ☐

c. we do not have a template (or set of instructions) for writing a SAP for an early phase clinical trial ☐

2i. If you selected option b, does this template have a specific early phase section?

Yes ☐ No ☐

2ii. If you selected option c, please briefly explain why

3. Are there any scenarios of early phase trials where you would **not** produce a SAP?

Yes ☐ No ☐

3a. If yes, for which types of early phase trials would you produce or not produce a SAP for? (If you do not perform these studies at your CTU then please select N/A). Please tick all that apply.

	Would produce	Would not produce	N/A
Rule-based phase I trials	☐	☐	☐
Model-based phase I trials	☐	☐	☐
Model-assisted phase I trials	☐	☐	☐
Randomised phase I trials	☐	☐	☐
Single arm phase II trials	☐	☐	☐
Other (please specify)	_____		

4. What analyses is the produced SAP(s) pertinent for:

Final analyses	Yes ☐	No ☐
Interim analyses	Yes ☐	No ☐
DMC/TSCs	Yes ☐	No ☐
Other (please specify)	_____	

5a. Would you be willing to provide a copy of the template used when writing SAPs for early phase clinical trials to contribute to the project as part of a review of UK CTU practice?

Yes ☐ No ☐

If yes, please forward along at the same time as returning this survey.

5b. Would you be willing to provide an example of an early phase SAP to contribute to the project as part of a review of UK CTU practice?

Yes ☐ No ☐

If yes, please forward along at the same time as returning this survey.

6. Do you think there is a requirement for early phase SAP guidance?

Yes ☐ No ☐

7. If such guidance existed, is this something you feel you would use at your CTU?

Yes ☐ No ☐

As part of this project, we will be producing guidelines, piloting them and holding consensus meetings. If your CTU would be happy to be involved with this project and be contacted in collaboration with it, please nominate a contact and provide their details below.

Name _____

Email _____

If you have any additional comment about SAPs for early phase trials, please use the space provided.

We would like to thank you for the time taken in considering our request and completing this survey. The opinions and views of your CTU are important for this project, and we are grateful for your support.