



Exploratory trials in mental health: anything to learn from other disciplines?

Sven Trelle

Correspondence to: Dr Sven Trelle, CTU Bern, University of Bern, Finkenhubelweg 11, Bern 3012, Switzerland; sven.trelle@ctu.unibe.ch

ABSTRACT

Objective Confirmatory randomised controlled trials require solid justifications, especially with regard to whether the experimental intervention is promising. Such evidence is generated in exploratory trials. However, empirical evidence shows that the quality of such trials is still suboptimal. More generally, the development process of healthcare interventions and especially of drugs, remains inefficient. Over the past 10–20 years, a vast amount of methodological work has been published about exploratory trials. This overview introduces some of the concepts and recent developments in the field.

Methods A narrative approach was taken for this overview. This article focuses on study designs developed outside the mental health field to introduce concepts that might not be familiar to clinical researchers in psychiatry and psychology. Non-randomised and randomised exploratory trial designs are covered. The article ends with a brief discussion on pilot studies and their difference to exploratory studies.

Results Classical designs for exploratory trials such as Simon's two-stage design still have a role. However, randomised exploratory trials are probably more suitable for mental health interventions. Newer, more flexible designs such as multistage, multiarm trials or platform trials have the potential to improve the efficiency of exploratory and subsequently confirmatory experiments.

Conclusions Although often not directly applicable, borrowing (study) design ideas from other medical disciplines has the potential to improve exploratory trials in the mental health field. At the same time, more explicit use of study designs specifically designed for exploratory trials will help to improve the transparency of such trials.

INTRODUCTION

Randomised controlled trials undoubtedly provide the best evidence for causal effects of healthcare interventions, for example, their effectiveness.¹ These trials are experimental studies involving human beings. Moreover, most healthcare interventions have positive and negative effects. Before healthcare interventions can be tested in an experimental study, a clear rationale needs to be provided. The rationale for an experimental study depends on the objective of the study at hand and the available evidence so far. For pharmaceutical products, this is represented by the development plan which is usually separated into four phases (I to IV).² It should be noted that most healthcare interventions often, but not always, develop roughly analogously over time, especially medical devices as well as non-drug, non-device interventions. Although clinical trials are often labelled according to the development phase, for example, a phase II trial, this approach may be abandoned mainly because such labels do not provide information on the objective and design of a particular trial. Instead, trials might be categorised according to their primary objective (figure 1).

The focus of this article is to provide an overview of study designs that precede the classical confirmatory randomised-controlled trial presenting classical designs and recent developments. Since the difference between pilot studies and exploratory trials is often not unambiguous, I will first discuss (therapeutic) exploratory trials and then briefly describe the concept of pilot studies focusing on the differences between these two types of studies. It must be acknowledged that the current overview is not based on a systematic literature search of methodological articles; nor are the provided examples a representative selection. Rather, as will be shown below, approaches to trials that precede confirmatory trials vary across medical disciplines and the article should sensitise investigators to the question whether the approaches taken in other disciplines might also be useful in mental health. Others noted that approaches that are common or even standard in some medical fields are still underused in other research areas.³ The methodology for exploratory trials has often been developed for oncological trials and the examples provided are often but not always taken from there. Finally, the inefficiency of the development process of therapeutic interventions, especially for drugs and the need for more efficient approaches have been discussed recently, including also the mental health field.^{4 5}

DESIGNING A THERAPEUTIC EXPLORATORY TRIAL

As mentioned in the introduction, the idea for therapeutic exploratory trials preceding a randomised controlled trial is to provide a clear rationale for the conduct of such a trial. The design of such trials varies and seems to depend on the medical discipline. In disciplines such as cardiology or mental health, these trials are often randomised, placebo-controlled trials. In contrast to confirmatory trials, these trials are often limited in sample size, short term and measure surrogate outcomes with the idea that results may help to predict success in a confirmatory trial. For example, a recently published trial of cariprazine in patients with acute mania associated with bipolar I disorder measured the primary outcome (Young Mania Rating Scale) at 3 weeks.⁶ However, the rationale for the different design choices was rather vague, which seems not uncommon in the mental health field according to a systematic review of trials in schizophrenia.⁷ Very few of the reviewed trials provided a clear primary hypothesis and the majority had no rationale for the chosen sample size. In contrast, exploratory trials in oncology are mainly single-arm trials that allow for interim analysis, although randomised designs are more common nowadays given their advantages.⁸ In the next sections, I will introduce various designs for exploratory trials. What is common to all these designs is that the designs provide a clear hypothesis and rationale for the sample size and criteria on trial success.

Simon's classical two-stage design

The most commonly used design and, at the same time, the origin of almost all exploratory trial designs is the two stage design as originally described by Simon.⁹ Although quite old already and designs with better operating characteristics exist, the design remains attractive to researchers as evidenced by its citation rate: according to the Web of Science and as of 2016, the original article⁹ received more than 2000 citations overall with more than 100 citations every year since 2004. The main reason for this is probably its simplicity and ease of implementation.

Trials based on Simon's two-stage design consist of two parts. The design includes an early stopping rule after the first stage based on a test for futility, that is, the trial stops if chances to observe a prespecified size of treatment effect are too small. In its original form, the design uses a binary outcome, for example, response to treatment, and requires the researcher to specify null and alternative response

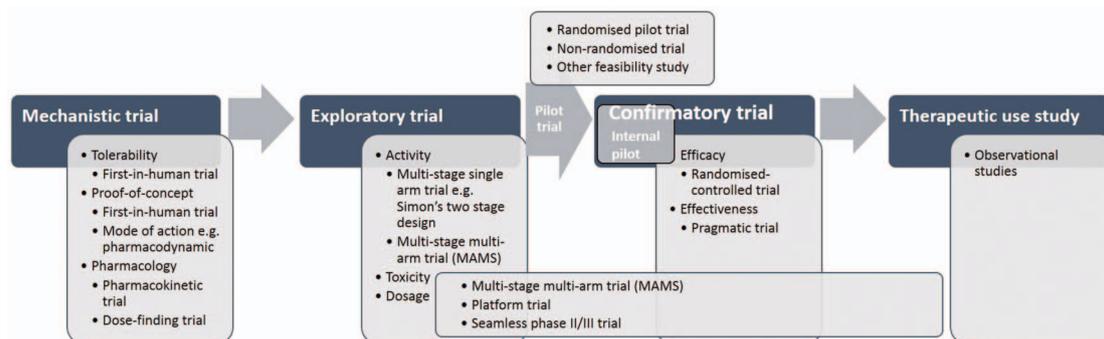


Figure 1 Development process of healthcare interventions (drugs) and related trials.

probabilities, that is, treatment response rates that are deemed clinically negligible and clinically important. Conceptually, this is closely related to a non-inferiority trial.¹⁰ It is important to realise that the determination of these probabilities is based on historical data, and therefore the design itself implicitly is one of historical comparisons and not concomitant comparisons as in a randomised trial.⁸ In addition, type I and II error probabilities, that is, α and power, need to be fixed. The design requires to recruit the calculated number of patients in the first stage. If the required number of responses is not observed at the end of stage one, the trial is stopped early; otherwise, it continues to the second stage. Success of the trial at the end of the second stage is determined based on whether the final observed response rate reaches the calculated cut-off.

The classical design has two variants to calculate the sample sizes for the different stages and the cut-offs for decision-making: one that minimises the expected sample size conditional on the null response probability (called optimal design) and one that minimises the total sample size (minimax design) dependent on the prespecified α and power. For example, assume that a response rate of 25% is deemed clinically too low and a response rate of 50% is deemed clinically interesting. Then, with α fixed at 0.05 and power at 90%, a trial will have to recruit 16 patients in the first stage. If four or fewer patients respond, the trial is stopped early; otherwise, it continues with an additional 17 patients for a total sample size of 33. If more than 12 patients out of these 33 responded, the trial is considered successful. The treatment is declared to be promising in the sense that it is likely to achieve the desirable 50% response rate. Sample size and cut-offs for decision-making are calculated iteratively. Routines are available in freely available software or websites as well as standard software packages such as STATA (simontwostage and simon2stage) or R (eg, clinfun, OneArmPhaseTwoStudy or ph2mult).

Variants of Simon's classical two-stage design

Numerous variants of the classical design have been developed since the original description in 1989 including improved algorithms to identify the best design, for example:

- ▶ Allowing for three^{11 12} or multiple stages¹³
- ▶ Allowing to monitor two binary outcomes such as response and a safety outcome¹⁴
- ▶ Allowing the use of continuous outcomes^{15 16}
- ▶ Allowing to stop early for futility and efficacy¹⁷

The lack of a comparator arm is an important caveat of all exploratory designs. This might not be an issue when testing an anticancer agent because spontaneous improvements/responses of patients are unlikely. In contrast, examining the efficacy of drugs in mental health conditions presents a challenge in the exploratory setting as high placebo response rates and spontaneous remissions are not unlikely. For example, a recent multiarm trial of different doses of lurasidone in patients with acute schizophrenia was considered to be a 'failed study' mainly because none

of the experimental treatment arms fared better than placebo and the response rate in the placebo group was considered to be unusually high.¹⁸ Moreover, designs described so far all come with the price inherent in non-concurrent comparisons, namely the difficulty in separating the inherent treatment effect of interest from trial effects such as patient selection, differences in cointerventions, other time effects or other confounders and biases.¹⁹ Nonetheless, the idea for stopping an exploratory trial early for futility may be attractive as it has the potential to reduce the number of patients who are unnecessarily included in a trial.³ Methods for monitoring of confirmatory trials²⁰ can be adapted to the exploratory setting if existing designs do not fit the individual needs. Since the variants of the original design are (much) more complex than Simon's original proposal, it is important to check the properties of the planned trial at the planning stage under different scenarios using simulation studies.

Randomised exploratory trial designs

To overcome the caveats, randomised exploratory designs have been proposed.²¹ Randomisation in these non-confirmatory trials can serve different purposes and three main approaches have been identified:¹⁹

1. Non-comparative approaches where randomised trial arms are considered separately as if they were single-arm trials
2. Comparative approaches where
 - A. Several experimental treatments are compared to select the most promising one for future trials and
 - B. Experimental treatment(s) are compared to standard of care or placebo to screen whether an experimental treatment is actual worth for evaluating it in a definitive confirmatory trial.

In the first approach, several individual single-arm trials are planned using the approaches described above, that is, different arms may or may not have a different sample size or different decision rules. Although patients are eventually randomised to the different arms, no comparison is made between arms but each arm is analysed independently. Such a design may be useful in situations where the lack of a comparator arm is of minor importance (eg, in oncology trials), but several experimental treatments are under consideration. Also, it is an attractive design for interventions early in their development phase where the available clinical evidence is very limited. If more than one experimental treatment meets the success criterion but only one or a few can be pursued for further evaluation, still no formal comparison is carried out between the treatments.²² Instead, comparisons are carried out informally by considering results on the primary outcome as well as secondary outcomes and then the winner is picked (see second approach). However, formal testing frameworks in such situations have also been described.²³

The second approach (2a) is often used when several variants of an experimental intervention are to be evaluated, for example, different dosages, schedules, etc. Patients are randomised to these different experimental arms, and these arms are eventually ranked with the aim

of selecting the most promising one (pick-the-winner design). Such a design was initially described by Simon, Wittes and Ellenberg (SWE) in 1985.²¹ However, the SWE design was shown to have unfavourable properties, especially with regard to the type I error rate.²⁴ Therefore, it is especially important to run simulation studies in the planning stage to see how the trial might run under different scenarios.

In the third approach (2b), patients are randomised to one or more experimental treatments and control with the aim of conducting non-definitive comparisons,²⁵ in the sense that the comparison results could be used to motivate but not replace a confirmatory trial. Therefore, the choice of the parameters that determine the sample size and potentially the success of the trial is critical. This concerns mainly the α level (type I error), the power and the effect size at which the trial is aiming. In particular, the α level will usually be higher than in a confirmatory setting. A recently published trial of the histamine H₃ receptor antagonist GSK239512 vs placebo for cognitive impairment in 80 stable patients with schizophrenia might be considered as a trial that falls into this category.²⁶ The rationale provided for the sample size of 80 patients was based on the width of the 90% CI and a target effect size. Another approach to such studies would be to use the concept of Simon's original two-stage design, namely the non-inferiority framework. For example, using the information of the previously mentioned GSK239512 trial, we can fix the following parameters: (1) target generic effect size/standardised mean difference of 0.4, (2) non-inferiority margin at an effect size of -0.42 , (3) α of 0.1. The power cannot be extracted from the article so we fix it at 80%. With these parameters, we arrive at a total sample size of 28, that is, less than half the sample size of the original trial. Changing the assumptions to a target effect size of 0.3 and a non-inferiority margin of -0.3 (usually considered a small effect), we arrive at a sample size of 52. Obviously, these parameters need to be adjusted to the individual trial at hand. Moreover, an interim analysis with a test for futility could easily be implemented.

Recently, the approaches 2a and 2b have gained new interest in the light of multiarm multistage (MAMS) clinical trials.²⁷ The methodology follows in principle the methodology of a standard randomised controlled trial but with additional adaptive elements²⁸ to improve the development process.²⁷ Strictly speaking, such trials are exploratory trials that should be considered what is usually called seamless phase II/III trials²⁸ and combine approaches 2a and b. Patients are randomised among several experimental treatments and control, for example, standard of care or placebo. Over time, experimental treatments are dropped and in a final stage, the most promising experimental treatment is compared to control. The selection or dropping of experimental trial arms over time can be achieved either by interim analyses or by adaptive randomisation.²⁹ Wason *et al*²⁷ provide a nice overview and also discuss applications for trials with continuous outcomes which are especially relevant for the mental health field.

Randomised exploratory trial designs can be fitted within a frequentist or Bayesian setting. In general, Bayesian designs allow for more flexibility. For example, it may be easier to conduct more interim analyses, incorporate more outcomes in such analyses, consider correlations across outcomes or add new treatment arms. This methodology is not often used as compared to frequentist approaches but recent advances in the methodology and related software suggest that Bayesian trial design might be more popular in the future.

Platform trials

Platform trials are typically implemented within a Bayesian framework, although a frequentist approach is also possible.^{4 30} A platform trial is an extension of an MAMS trial, but usually with broader goals and a

much more flexible trial protocol. The main aim is to find the best treatment option(s). The main difference to the trials discussed so far is that a platform trial runs long term, that is, potentially for an indefinite amount of time or as long as there are suitable treatments under evaluation because trial arms are dropped and added continuously.⁴ Decision rules guide the process of adding or dropping treatment arms, changes in eligibility criteria, for example, dropping patient subgroups, or combining treatments in a new arm. To the best of my knowledge, only one trial, in patients with breast cancer, has published first results using such a design.^{31 32} However, others are underway including a trial in Alzheimer's disease.³³

A primer on pilot studies (or 'proof of concept' trials)

Pilot studies have long been neglected in clinical research. However, such studies have recently gained attention as evidenced by a recently published extension to the CONSORT statement³⁴ or the emergence of a peer-reviewed journal dedicated to the publication of such studies (<https://pilotfeasibilitystudies.biomedcentral.com/>). A conceptual framework to distinguish pilot studies from other types of studies is also now available.³⁵ Within this framework, pilot studies evaluate parts or the complete planned confirmatory trial. The idea of these studies is not to test whether a treatment is promising for confirmatory testing in a randomised controlled trial as it is in a (therapeutic) exploratory trial. Rather, such studies evaluate the feasibility of the study design and directly precede a confirmatory trial, if successful.³⁶ Common domains to be evaluated in a pilot study are:

- ▶ Recruitment potential
- ▶ Recruitment, informed consent or randomisation process
- ▶ Assessments and outcome measures
- ▶ Data collection and follow-up
- ▶ Resource use and costs
- ▶ Collecting outcome data to inform the sample size calculation of the main trial

As such, these studies can be quantitative as well as qualitative and may be conducted separately from the main trial or nested within the main trial (internal pilot). Although some internal piloting is common to most randomised controlled trials,³⁷ it should be stressed that these are usually not formally defined but conducted rather informally. In contrast, a true pilot study has a study protocol and prespecified outcomes, although some flexibility is inevitable and actually desirable.³⁸ Retrospective analyses of trials that were discontinued prematurely suggest that performing a full pilot study before the actual trial is protective against premature discontinuation, especially with regard to poor recruitment.³⁹ Their usefulness is therefore generally not challenged. Thus, whether to conduct a pilot or not is mainly determined by available time and resources.

CONCLUSIONS

Over the past 10–20 years, a rich body of methodological work has been published on study designs for therapeutic exploratory trials. Designs for almost all situations are available. It is therefore surprising that the quality of such trials is still poor including an unspecified hypothesis, unclear objectives and non-existing sample size justifications.⁷ One reason may be that investigators feel uncomfortable in borrowing ideas from other medical disciplines or feel that these approaches are not applicable to their situation. However, from my experience, similarities usually outweigh the differences. Moreover, most designs are flexible and can be adapted to fit individual needs. It is therefore hoped that we will see more efficient development of healthcare interventions in the near future with more transparent and, at the same time, comprehensive and flexible trials.

Competing interests None declared.

Provenance and peer review Not commissioned; externally peer reviewed.

doi:10.1136/eb-2016-102581

Received 15 December 2016; Accepted 16 December 2016

REFERENCES

- Hernan MA**, Robins JM. *Causal inference*. Cambridge, MA, 2015. <https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/> (accessed 25 Jul 2016).
- International Conference on Harmonisation**. *ICH Harmonised Tripartite Guideline: general considerations for clinical trials E8*. London, UK: International Conference on Harmonisation, 1997.
- Palesch YY**, Tilley BC, Sackett DL, *et al*. Applying a phase II futility study design to therapeutic stroke trials. *Stroke* 2005;**36**:2410–14.
- Berry SM**, Connor JT, Lewis RJ. The platform trial: an efficient strategy for evaluating multiple treatments. *JAMA* 2015;**313**:1619–20.
- Cummings JL**, Morstorf T, Zhong K. Alzheimer's disease drug-development pipeline: few candidates, frequent failures. *Alzheimers Res Ther* 2014;**6**:37.
- Durgam S**, Starace A, Li D, *et al*. The efficacy and tolerability of cariprazine in acute mania associated with bipolar I disorder: a phase II trial. *Bipolar Disord* 2015;**17**:63–75.
- Patel MX**, Collins S, Hellier J, *et al*. The quality of reporting of phase II and III trials for new antipsychotics: a systematic review. *Psychol Med* 2015;**45**:467–79.
- Sharma MR**, Stadler WM, Ratain MJ. Randomized phase II trials: a long-term investment with promising returns. *J Natl Cancer Inst* 2011;**103**:1093–100.
- Simon R**. Optimal two-stage designs for phase II clinical trials. *Control Clin Trials* 1989;**10**:1–10.
- Schumi J**, Wittes JT. Through the looking glass: understanding non-inferiority. *Trials* 2011;**12**:106.
- Ensign LG**, Gehan EA, Kamen DS, *et al*. An optimal three-stage design for phase II clinical trials. *Stat Med* 1994;**13**:1727–36.
- Chen TT**. Optimal three-stage designs for phase II cancer clinical trials. *Stat Med* 1997;**16**:2701–11.
- Chen S**, Soong SJ, Wheeler RH. An efficient multiple-stage procedure for phase II clinical trials that have high response rate objectives. *Control Clin Trials* 1994;**15**:277–83.
- Bryant J**, Day R. Incorporating toxicity considerations into the design of two-stage phase II clinical trials. *Biometrics* 1995;**51**:1372–83.
- Wason JM**, Mander AP, Eisen TG. Reducing sample sizes in two-stage phase II cancer trials by using continuous tumour shrinkage end-points. *Eur J Cancer* 2011;**47**:983–9.
- Wason JM**, Mander AP, Thompson SG. Optimal multistage designs for randomised clinical trials with continuous outcomes. *Stat Med* 2012;**31**:301–12.
- Mander AP**, Thompson SG. Two-stage designs optimal under the alternative hypothesis for phase II cancer clinical trials. *Contemp Clin Trials* 2010;**31**:572–8.
- Potkin SG**, Kimura T, Guarino J. A 6-week, double-blind, placebo- and haloperidol-controlled, phase II study of lurasidone in patients with acute schizophrenia. *Ther Adv Psychopharmacol* 2015;**5**:322–31.
- Mandrekar SJ**, Sargent DJ. Randomized phase II trials: time for a new era in clinical trial design. *J Thorac Oncol* 2010;**5**:932–4.
- Lachin JM**. A review of methods for futility stopping based on conditional power. *Stat Med* 2005;**24**:2747–64.
- Simon R**, Wittes RE, Ellenberg SS. Randomized phase II clinical trials. *Cancer Treat Rep* 1985;**69**:1375–81.
- Wieand HS**. Randomized phase II trials: what does randomization gain? *J Clin Oncol* 2005;**23**:1794–5.
- Jung SH**, George SL. Between-arm comparisons in randomized phase II trials. *J Biopharm Stat* 2009;**19**:456–68.
- Liu PY**, LeBlanc M, Desai M. False positive rates of randomized phase II designs. *Control Clin Trials* 1999;**20**:343–52.
- Rubinstein LV**, Korn EL, Freidlin B, *et al*. Design issues of randomized phase II trials and a proposal for phase II screening trials. *J Clin Oncol* 2005;**23**:7199–206.
- Jarskog LF**, Lowy MT, Grove RA, *et al*. A phase II study of a histamine H(3) receptor antagonist GSK239512 for cognitive impairment in stable schizophrenia subjects on antipsychotic therapy. *Schizophr Res* 2015;**164**:136–42.
- Wason J**, Magirr D, Law M, *et al*. Some recommendations for multi-arm multi-stage trials. *Stat Methods Med Res* 2016;**25**:716–27.
- Chow SC**, Chang M. Adaptive design methods in clinical trials—a review. *Orphanet J Rare Dis* 2008;**3**:11.
- Wason JM**, Trippa L. A comparison of Bayesian adaptive randomization and multi-stage designs for multi-arm clinical trials. *Stat Med* 2014;**33**:2206–21.
- Wason JM**, Jaki T. A review of statistical designs for improving the efficiency of phase II studies in oncology. *Stat Methods Med Res* 2016;**25**:1010–21.
- Park JW**, Liu MC, Yee D, *et al*. Adaptive randomization of neratinib in early breast cancer. *N Engl J Med* 2016;**375**:11–22.
- Rugo HS**, Olopade OI, DeMichele A, *et al*. Adaptive randomization of veliparib-carboplatin treatment in breast cancer. *N Engl J Med* 2016;**375**:23–34.
- Bateman RJ**, Benzinger TL, Berry S, *et al*. The DIAN-TU Next Generation Alzheimer's prevention trial: adaptive design and disease progression model. *Alzheimers Dement* 2017;**13**:8–19.
- Eldridge SM**, Chan CL, Campbell MJ, *et al*. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. *BMJ* 2016;**355**:i5239.
- Eldridge SM**, Lancaster GA, Campbell MJ, *et al*. Defining feasibility and pilot studies in preparation for randomised controlled trials: development of a conceptual framework. *PLoS ONE* 2016;**11**:e0150205.
- Araim M**, Campbell MJ, Cooper CL, *et al*. What is a pilot or feasibility study? A review of current practice and editorial policy. *BMC Med Res Methodol* 2010;**10**:67.
- Cook JA**, Beard DJ, Cook JR, *et al*. The curious case of an internal pilot in a multicentre randomised trial—time for a rethink? *Pilot Feasibility Stud* 2016;**2**:73.
- Thabane L**, Ma J, Chu R, *et al*. A tutorial on pilot studies: the what, why and how. *BMC Med Res Methodol* 2010;**10**:1.
- Kasenda B**, von Elm E, You J, *et al*. Prevalence, characteristics, and publication of discontinued randomized trials. *JAMA* 2014;**311**:1045–51.