Expert consensus recommendations on the use of randomized clinical trials for drug approval in psychiatry- comparing trial designs

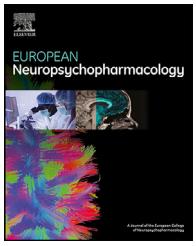
Similon, Miriam von Mucke; Paasche, Cecilia; Krol, Fas; Lerer, Bernard; Goodwin, Guy M.; Berk, Michael; Meyer-Lindenberg, Andreas; Ketter, Terence A.; Yatham, Lakshmi N.; Goldberg, Joseph F.

Published in:
European Neuropsychopharmacology

DOI:
10.1016/j.euroneuro.2022.05.002

Publication date:
2022

European Neuropsychopharmacology 60 (2022) 91–99



www.elsevier.com/locate/euroneuro

Expert consensus recommendations on the use of randomized clinical trials for drug approval in psychiatry- comparing trial designs

Miriam von Mücke Similon[a,b,1], Cecilia Paasche[a,b,1], Fas Krol[c], Bernard Lerer[d], Guy M. Goodwin[e], Michael Berk[f], Andreas Meyer-Lindenberg[g], Terence A. Ketter[h], Lakshmi N. Yatham[i], Joseph F. Goldberg[j], Gin S. Malhi[k1,k2,k3], Rif El-Mallakh[l], Rasmus W. Licht[m], Allan H. Young[n1,n2], Flavio Kapczinski[o], Marnina Swartz[a], Michal Hagin[a], Carla Torrent[p], Alessandro Serretti[q], Ayşegül Yildiz[r], Anabel Martínez-Arán[p], Sergio Strejilevich[s], Janusz Rybakowski[t], Gabriele Sani[u1,ub], Heinz Grunze[v], Gustavo Vázquez[w], Ana Gonzales Pinto[x], Jean Michel Azorin[y], Willem Nolen[z], Othman Sentissi[aa], Carlos López-Jaramillo[bb], Benicio N. Frey[j,cc], Andrew Nierenberg[dd], Gordon Parker[ee], David J Bond[ff], Adam Cohen[gg], Alfonso Tortorella[hh], Giulio Perugi[ii], Eduard Vieta[p], Dina Popovic[a,*]

[a] *Abarbanel Mental Health Center, Bat Yam, Israel*
[b] *Tel Aviv University, Sackler School of Medicine, Israel*
[c] *Leiden University Medical Center, the Netherlands*
[d] *Hadassah Medical University, Jerusalem, Israel*
[e] *Department of Psychiatry, University of Oxford, Oxford, UK*
[f] *Deakin University, IMPACT – the Institute for Mental and Physical Health and Clinical Translation, School of Medicine, Barwon Health, Geelong, Australia*
[g] *Department of Psychiatry and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim / University of Heidelberg, Mannheim, Germany*

[*] Corresponding author at: Abarbanel Mental Health Center, Keren-Kayemet Street 15, Bat Yam, Israel
E-mail address: Popovic.dina@gmail.com (D. Popovic).
[1] Both contributed equally to the writing of this paper

https://doi.org/10.1016/j.euroneuro.2022.05.002

[h] *Department of Psychiatry and Behavioral Sciences Stanford, University School of Medicine Stanford, California, USA*
[i] *Department of Psychiatry, University of British Columbia, Vancouver, British Columbia, Canada*
[j] *Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, New York*
[k1] *The University of Sydney, Faculty of Medicine and Health, Northern Clinical School, Department of Psychiatry, Sydney, New South Wales, Australia*
[k2] *Academic Department of Psychiatry, Royal North Shore Hospital, Northern Sydney Local Health District, St Leonards, NSW 2065 Australia*
[k3] *CADE Clinic, Royal North Shore Hospital, Northern Sydney Local Health District, St Leonards, NSW 2065 Australia*
[l] *University of Louisville School of Medicine, Louisville, Kentucky, USA*
[m] *Aalborg University Hospital, Psychiatry, Aalborg Denmark and Clinical Department of Medicine, Aalborg University, Aalborg, Denmark*
[n1] *Department of Psychological Medicine, Institute of Psychiatry, Psychology and Neuroscience, King's College London & South London and Maudsley NHS Foundation Trust, Bethlem Royal Hospital, Monks Orchard Road, Beckenham, Kent, BR3 3BX*
[n2] *United Kingdom. Department of Psychiatry, Harvard Medical School, and the Bipolar Clinic and Research Program, Massachusetts General Hospital, Boston, USA*
[o] *Department of Psychiatry and Behavioural Neurosciences, McMaster University, Hamilton, ON, Canada*
[p] *Bipolar Disorders Program, Bipolar Disorders Unit, Hospital Clinic, University of Barcelona, IDIBAPS, CIBERSAM, Barcelona, Catalonia, Spain*
[q] *Department of Biomedical and Neuro Motor Sciences, University of Bologna, Bologna, Italy*
[r] *Department of Psychiatry, Dokuz Eylül University, Izmir, Turkey*
[s] *Assistance and Research in Affective Disorders, Buenos Aires, Argentina*
[t] *Department of Adult Psychiatry, Poznan University of Medical Sciences, Poland*
[u1] *Department of Neuroscience, Section of Psychiatry, Università Cattolica del Sacro Cuore, Rome, Italy*
[ub] *Department of Psychiatry, Fondazione Policlinico Universitario Agostino Gemelli IRCCS, Rome, Italy*
[v] *Psychiatrie Schwäbisch Hall & Paracelsus Medical University, Nuremberg, Germany*
[w] *Department of Psychiatry, School of Medicine, Queen's University, Ontario, Canada*
[x] *Department of Psychiatry, BIOARABA. Araba University Hospital, University of the Basque Country, Vitoria, Spain*
[y] *Hôpital Sainte-Marguerite, Marseille, France*
[z] *Psychiatrie, Universitair Medisch Centrum Groningen, the Netherlands*
[aa] *Division of Adult Psychiatry, Department of Psychiatry, Geneva University Hospital, 2, Chemin du Petit-Bel-Air, CH-1226 Thonex, Switzerland*
[bb] *Universidad de Antioquia, Hospital Universitario San Vicente Fundación, Columbia*
[cc] *Mood Disorders Program and Women's Health Concerns Clinic, St. Joseph's Healthcare Hamilton, ON, Canada*
[dd] *Bipolar Clinic and Research Program, Massachusetts General Hospital, Boston, MA, USA*
[ee] *School of Psychiatry, University of New South Wales, Australia*
[ff] *Department of Psychiatry and Behavioral Sciences, University of Minnesota, Minneapolis, MN, USA*
[gg] *Leiden University Medical Center, the Netherlands*
[hh] *Department of Psychiatry, University of Perugia, Italy*
[ii] *University of Pisa, Italy*

Received 12 August 2021; received in revised form 2 May 2022; accepted 4 May 2022

KEYWORDS
Placebo;
Delphi;
Clinical trial;
Trial-design

Abstract
The use of randomized clinical trials, in particular placebo-controlled trials, for drug approval, is the subject of long-standing debate in the scientific community and beyond. This study offers consensus recommendations from clinical and academic experts to guide the selection of clinical trial design in psychiatry. Forty-one highly cited clinical psychiatrists and/or researchers participated in a Delphi survey. Consensus statements were developed based on the findings of a published, peer-reviewed systematic review. Participants evaluated statements in two survey rounds, following the Delphi method. The expert panel achieved consensus on 7 of 21 recommendations regarding the use of randomized clinical trials. The endorsed recommendations were: (i) Results from placebo-controlled trials are the most reliable and (ii) are necessary despite the growing placebo-effect; (iii) it is ethical to enroll patients in placebo-arms when established treatment is available, if there is no evidence of increased health risk; (iv) There is a need to approve new drugs with the same efficacy as existing treatments, but with different side-effect profiles; (v) Non-inferiority trials incur an increased risk of approving ineffective medications; (vi) The risk of approving an ineffective drug justifies trial designs that incur higher costs, and (vii) superiority trials incur the risk of rejecting potentially efficacious treatments. The endorsed recommendations inform the choice of trial-design appropriate for approval of psychopharmacological drugs. The recommendations strongly support the use of randomized clinical trials in general, and the use of placebo-controlled trials in particular.

1. Introduction

In modern psychiatry, and medicine in general, finding the appropriate experimental design that provides reliable signal detection for medication efficacy and ensures trial integrity is of paramount importance. Randomized clinical trials (RCTs) include placebo-controlled and active-comparator trials. Placebo-controlled trials are widely considered the gold standard for the evaluation of new medications (Healy, 2003; Vieta & Cruz, 2012). In a placebo-controlled trial, the efficacy of the experimental intervention is established by demonstration of its superiority to placebo in producing the desired effect (Krol et al., 2020). Trials involving active comparators can be either superiority trials or non-inferiority trials. In a superiority trial, a new agent is compared to an existing agent(s) to evaluate if the new one outperforms the old in terms of efficacy and/or tolerability. A non-inferiority trial also compares a new agent against an established one, but the new agent is required to demonstrate that it is not worse than the proven drug (Pocock, 2003).

The use of placebo raises several issues that are the subject of uncertainty and informal debate. First, placebo is not a standard of care, and placebo tablets per se are not part of modern medical practice. Placebo controlled randomized trials are accordingly experiments, always somewhat removed from ordinary care. Second, rates of 'response' to placebo appear to have been rising in some psychiatric indications, notably in major depression disorder (Howick, 2009; Khan et al., 2017; Temple & Ellenberg, 2000; Walsh et al., 2002). Placebo, in the context of a clinical trial, is both a physical object (a pharmacologically inert pill, identical in appearance to an active comparator) but also probably a psychosocial intervention related to the intensity of attention and care in a clinical trial (de Craen et al., 1999; Ovosi et al., 2017). This usually exceeds the attention granted in ordinary practice. Such non-specific effects can magnify the placebo effect; it can then become impossible to detect the superiority of an active comparator. In addition, it can be difficult to distinguish negative trials from failed trials when unexpectedly high placebo response rates eclipse the potential intrinsic efficacy of a pharmacologic compound, essentially increasing the risk of a Type II error (Kemp et al., 2010; Pocock & Stone, 2016). Finally, patient groups must have assay sensitivity; the experimental design must have potential to show a differential effect of treatment by appropriate choice of treatment groups (Howick, 2009). In the extreme case, some patients may respond to any intervention and some may respond to none: inclusion of large numbers of either such patient will invalidate any trial design.

The choice of trial design has an enormous impact on the outcomes of a study, and the choice is influenced by different factors, that are not necessarily evidence-based. Based on the findings of a systematic review of the literature by Krol et al. (2020), a Delphi study (Linstone & Turoff, 1975) was conducted to obtain consensus among expert clinicians and researchers on the role of RCTs in clinical psychopharmacology. Defining current clinical opinion could help to resolve any discrepancy between what is allowed by the declaration of Helsinki, the set of ethical principles developed by the World Medical Association for conducting human experiments, and what researchers feel comfortable performing (Batra & Howick, 2017).

2. Experimental procedures

2.1. Selection of statements for the Delphi survey

Following a published systematic review of the relevant literature, 21 value statements were chosen based on prevalent arguments regarding different types of RCTs (superiority, non-inferiority, placebo-controlled) (Krol et al., 2020; E. Vieta & Cruz, 2012). An expert task force iteratively developed consensus through serial consensus-based revisions using the Delphi method. Following the survey of initial items, subsequent survey included items that needed to be re-rated. This process resulted in the final clinical recommendations.

2.2. Consensus Method

2.2.1. EXPERT PANEL

The expert task force was composed of a panel of international experts on research in psychiatry, selected according to an objective procedure based on a Scopus search of citations on the specific topic of research in mood disorders, psychotic disorders, and psychopharmacology.

The most cited authors and some additional authors from key geographical areas were identified and invited by e-mail to participate; 53.2% (41/77) agreed to participate. Experts from 17 countries in Europe, North America, South America, Australia and Middle East Asia participated. The steering committee (D.P., F.K., B.L., E.V., and M.H.) initially focused on the discussion and integration of findings from peer-reviewed published research findings on the topic (Krol et al., 2020). Based on the findings of the review, queries were prepared and approved by the steering committee.

2.2.2. Delphi method

The Delphi method is an iterative process of asking an expert panel for their agreement with certain statements (Linstone & Turoff, 1975). It was first utilized to predict defense technology during the Cold War (Dalkey & Helmer, 1963), but has now been applied to the social sciences and medicine, including psychiatry (Lintonen et al., 2014; Nolen et al., 2019; Pacchiarotti et al., 2013; Popovic et al., 2014).

Following a systematic review of the literature, the Steering Committee formulated 21 statements regarding the use of placebo, superiority, and non-inferiority trial designs in psychiatry. This formal process used an anonymous iterative survey series wherein participants rated statements based on their agreement, using a five-point scale ranging from '*strongly disagree*' to '*strongly agree*.' Survey participants did not know the other participants' responses in each round. After each of the two rounds, the level of conformity was calculated by grouping the answers into '*agree*,' '*neutral*,' or '*disagree*.'

Criteria for the Delphi method were established *a priori* based on similar surveys (Pacchiarotti et al., 2013; Popovic et al., 2015). *Endorsed items* were the items rated as essential or important by at least 80% of participants in both rounds and were included in the recommendations. *Re-rated items:* Items rated as essential or important by 65-79% of panel experts in the first round, that were then re-rated in round 2. *Rejected items* were items that received a consensus rating of <65% in round 1, or re-rated items that achieved <80% rating in round 2 (Table 3). Final recommendations were those that achieved at least an 80% rating in round 1 or round 2 (Table 1).

3. Results

The first survey round was rated by 41 participants and 33 responded to both rounds (Fig. 1). A total of 7 statements derived from the original 21 survey items were endorsed over the two survey rounds by 80% or more of the expert panel (Table 1). These consensus statements form the recommendations for the choice of experimental design in randomized clinical trials. The endorsed recommendations were: (i) Results from placebo-controlled trials are the most reliable and (ii) are necessary despite the growing placebo-effect; (iii) it is ethical to enroll patients in placebo-arms when established treatment is available, if there is no evidence of increased health risk; (iv) There is a need to approve new drugs with the same efficacy as existing treatments, but with different side-effect profiles; (v) Non-inferiority trials incur an increased risk of approving ineffective medications; (vi) The risk of approving an ineffective drug justifies trial designs that incur higher costs, and (vii) superiority trials incur the risk of rejecting potentially efficacious treatments. The 14 statements that did not achieve consensus over 80% are shown in Table 3.

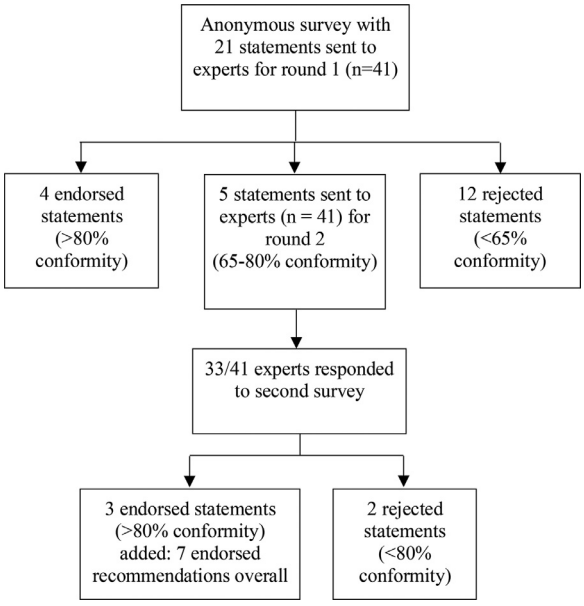


Fig. 1 Flow diagram of Delphi study design and results

4. Discussion

The results of the present study yielded seven endorsed recommendations on the choice of trial-design in psychiatry (Table 1). Experts agreed that results from placebo-controlled trials are more reliable than results from any other study-design, and that the growing placebo-effect does not negate the need for placebo-controlled trials. Regarding non-inferiority trials, the panel concluded that the increased chance of approving ineffective medications in non-inferiority trials should impact the choice of trial design. Moreover, they agreed that the risk of approving an ineffective drug justifies trial designs that incur higher costs. The experts concurred that use of an active comparator instead of placebo leads to a risk of rejecting valuable new treatments, because it is more difficult to prove superiority of a medication in active comparator trials. Along these lines, the panel agreed that it is justified to approve a new

Table 1 Recommendations endorsed by ≥ 80% conformity

Endorsed Recommendations	Conformity (%)
1. It is justified to approve a new drug with the same efficacy as existing treatments, but with a different side-effect profile.	100 (round 1)
2. The risk of approving an ineffective drug justifies trial designs that incur higher costs (more subjects, longer duration etc.).	85.4 (round 1)
3. Results from placebo-controlled trials are more reliable than results from any other study design.	85.3 (round 1)
4. The growing placebo-effect does not negate the need for placebo-controlled trials (vs. only using active-controlled trials).	83 (round 1)
5. It is not unethical to enroll patients in a placebo-arm when established treatment is available, if there is no evidence of increased health risk.	75.6 (round 1) 87.9 (round 2)
6. The increased chance of approving ineffective medications in non-inferiority trials should impact the choice of trial design.	70.8 (round 1) 81.8 (round 2)
7. Use of an active comparator instead of placebo leads to a risk of rejecting valuable new treatments, because it is more difficult to prove superiority of a medication in an active comparator trial.	70.7 (round 1) 84.9 (round 2)

drug with the same efficacy as existing treatments, but with a different side-effect profile. With respect to the ethics of trial-design, there was consensus among the panelists that it is *not* unethical to enroll patients in a placebo-arm when established treatment is available, if there is no evidence of increased health risk.

4.1. Placebo-controlled trials

The Declaration of Helsinki in 1996 from the World Medical Association initially stated that patients should always receive the best-known therapy (WMA, 1996); this was often simply ignored when choosing trial designs (Glass, 2008; Levine, 1999). The declaration has had subsequent amendments; the latest version appeared in 2018 (WMA, 2018) and states in relation to placebo:

Where no proven intervention exists, the use of placebo, or no intervention, is acceptable; or Where for compelling and scientifically sound methodological reasons the use of any intervention less effective than the best proven one, the use of placebo, or no intervention is necessary to determine the efficacy or safety of an intervention and the patients who receive any intervention less effective than the best proven one, placebo, or no intervention will not be subject to additional risks of serious or irreversible harm as a result of not receiving the best proven intervention.

Interpretations of these qualifications clearly remain open to debate (Batra & Howick, 2017). Thus, how should we define 'compelling and scientifically sound reasons' or understand what harms are acceptable (and not serious or irreversible).

In the present study experts agreed that "results from placebo-controlled trials are more reliable than results from any other study design" (Table 1, Table 2). One argument for the reliability of placebo-controlled trials in the literature states and re-states that the design allows for a differential effect of treatment to be demonstrated, if one exists (D'Agostino et al., 2003; Ellenberg & Temple, 2000; Garattini & Bertele, 2009; Krol et al., 2020; Powers et al., 2005; Snapinn, 2014; Spławiński & Kuźniar, 2004; Streiner, 2008; Temple & Ellenberg, 2000; Vieta & Cruz, 2012).

There may also be a 7% increase per decade in the placebo effect in trial for major depression (Walsh et al., 2002). The expert panel agreed that "the growing placebo-effect does not negate the need for placebo-controlled trials" (vs only using active-controlled trials) (Table 1). However, the foregoing data suggest that caution may be needed in accepting this opinion for major depression. A rising placebo response and a reduced active drug effect in placebo-controlled trials risks rejecting effective new treatments for depression.

4.2. Non-inferiority trials

The corollary to the statement that placebo-controlled trials are the most reliable, is that experts consider non-inferiority and superiority trials less reliable. The argument that non-inferiority trials are less statistically sound is based on the idea that they are too permissive (vs. superiority or placebo-controlled trials) and yield approval of weak drugs with questionable benefit-risk ratios (D'Agostino et al., 2003; Powers et al., 2005; Snapinn, 2014; Spławiński & Kuźniar, 2004; Streiner, 2008; Temple & Ellenberg, 2000; Vieta & Cruz, 2012). The expert panel considers that the increased chance of approving ineffective medications in non-inferiority trials *should* impact the choice of trial design (Table 1). A valid non-inferiority trial cannot assume that the active comparator will always show effects superior to placebo; this is not always the case. High placebo or non-specific response rates may occur across both active arms (D'Agostino et al., 2003; Ellenberg & Temple, 2000; Powers et al., 2005; Snapinn, 2014; Spławiński & Kuźniar, 2004; Streiner, 2008; Temple & Ellenberg, 2000; Vieta & Cruz, 2012). Non-inferiority trials need great care because an error-ridden trial will show equivalence between treatments when there is none, driving the publication of invalid results (Ellenberg & Temple, 2000; Garattini & Bertele, 2009; Snapinn, 2014; Spławiński & Kuźniar, 2004; Streiner, 2008; Temple & Ellenberg, 2000;

Table 2 Literature background of endorsed recommendations

Recommendations	Papers that cite argument
1. It is justified to approve a new drug with the same efficacy as existing treatments, but with a different side-effect profile.	Batra & Howick, 2017; Burger et al., 2011; D'Agostino et al., 2003; Ellenberg & Temple, 2000; Fleischhacker et al., 2003; Garattini & Bertele, 2009; Howick, 2009; Pocock, 2003; Shapiro et al., 2010; Vieta & Cruz, 2012
2. The risk of approving an ineffective drug justifies trial designs that incur higher costs (more subjects, longer duration etc.)*	D'Agostino et al., 2003; Splawinski & Kuzniar, 2004; Vieta & Cruz, 2012
3. Results from placebo-controlled trials are more reliable than results from any other study design.	D'Agostino et al., 2003; Fleischhacker et al., 2003; Garattini & Bertele, 2009; Hasnain et al., 2018; Ovosi et al., 2017; Splawinski & Kuzniar, 2004; Streiner, 1995; Temple & Ellenberg, 2000; Vieta & Cruz, 2012; Walsh et al., 2002
4. The growing placebo-effect does not negate the need for placebo-controlled trials (vs. only using active-controlled trials).	D'Agostino et al., 2003; Hasnain et al., 2018; Pocock, 2003; Snapinn, 2014; Splawinski & Kuzniar, 2004; Streiner, 2008; Vieta & Cruz, 2012; Walsh et al., 2002
5. It is not unethical to enroll patients in a placebo- arm when established treatment is available, if there is no evidence of increased health risk.	Ellenberg & Temple, 2000; Hasnain et al., 2018; Temple & Ellenberg, 2000; Vieta & Cruz, 2012; Walsh et al., 2002
6. The increased chance of approving ineffective medications in non-inferiority trials should impact the choice of trial design.	Burger et al., 2011; Powers et al., 2005; Snapinn, 2014; Splawinski & Kuzniar, 2004; Temple & Ellenberg, 2000; Vieta & Cruz, 2012
7. Use of an active comparator instead of placebo leads to a risk of rejecting valuable new treatments, because it is more difficult to prove superiority of a medication in an active comparator trial.	Ellenberg & Temple, 2000; Pocock, 2003; Vieta & Cruz, 2012

*= Arguments in favor of superiority trials, which are the most expensive trials

Table 3 Statements that did not reach consensus

Statements	Conformity (%)
1. Researchers should not always prefer an active-controlled superiority trial over a placebo-controlled trial.	66.7 (round 2)
2. Approval of new drugs should always require a placebo-controlled trial.	66.7 (round 2)
3. When treating my patients, I feel comfortable prescribing drugs for off-label indications.	61.0 (round 1)
4. Results from a non-inferiority trial are not as reliable as results from a placebo-controlled trial.	58.5 (round 1)
5. Despite the larger sample size needed in active-controlled superiority trials vs. placebo-controlled trials, it is not justified to expose more patients to a trial drug with unproven efficacy, than expose a lower number of subjects to placebo.	53.7 (round 1)
6. In clinical practice, I would prescribe a new drug that wasn't tested against placebo.	51.2 (round 1)
7. Considering the increase in placebo response over the years, results from older studies can no longer be directly compared to results from recent studies.	51.2 (round 1)
8. Easier recruitment of participants for active-controlled superiority or non inferiority trials compensates for the larger sample sizes than in placebo- controlled trials.	48.7 (round 1)
9. Established treatments used as active comparators give us less information than placebo.	46.4 (round 1)
10. FDA and EMA should not approve drugs that have not been tested against placebo.	46.4 (round 1)
11. It is justifiable to suspend a patient's ongoing treatment in a placebo- controlled trial in order to obtain more reliable data on new treatments	46.4 (round 1)
12. Calculated measures of effect size, such as NNT, should not be used as an alternative to head-to-head trials to compare efficacy of different drugs.	46.4 (round 1)
13. The risk of biocreep, the gradual degradation of efficacy of newly approved treatments, does not affect the decision to conduct non-inferiority trials.	46.3 (round 1)
14. Results of active-controlled superiority trials are less reliable than results of placebo-controlled trials.	43.9 (round 1)

Vieta & Cruz, 2012). In principle, a third placebo arm (not necessarily powered for significance) can offer reassurance that a non-inferiority trial is valid.

4.3. Superiority trials

The literature shows disadvantages in convenience and cost, yet advantages in reliability compared to non-inferiority trials (Batra & Howick, 2017; Ellenberg & Temple, 2000; Spławiński & Kuźniar, 2004; Temple & Ellenberg, 2000; Vieta & Cruz, 2012). The expert panelists agreed upon practical disadvantages of superiority trials that extend beyond convenience and cost; they could lead to the rejection of valuable new treatments, because it is more difficult to prove superiority of a medication in an active comparator trial. (Table 1). Statistically it may be hard to prove superiority of one efficacious drug over another (Temple & Ellenberg, 2000; Vieta & Cruz, 2012). The effect size difference between two active arms is likely to be smaller than that between an active and placebo arm. This may markedly increase the sample size requirements. If this were required, as some authorities suggest (Barbui & Bighelli, 2013; Spielmans & Kirsch, 2014), it would likely lead to fewer drug approvals and fewer treatment options for patients. It is advantageous to have different drugs with the same indication due to differences in individual tolerability, in drug-drug interactions, and in potential efficacy for certain subgroups of patients, in the context of stratified treatment to optimize care for each patient (Murawiec & Popovic, 2015; Pocock, 2003; Popovic et al., 2012; Vieta & Cruz, 2012). There is an overwhelming consensus (i.e. 100%) among experts with the statement that, "It is justified to approve a new drug with the same efficacy as existing treatments, but with a different side-effect profile" (Table 2), which is in alignment with the literature (Fleischhacker et al., 2003; Pocock, 2003; Vieta & Cruz, 2012). The heterogeneity of psychiatric disorders leads to variable drug response between patients, so drugs with similar effect sizes may have distinct clinical impact (Pocock, 2003; Vieta & Cruz, 2012).

4.3.1. Limitations of RCTs

RCTs are used mainly for regulatory purposes in new industry pharmaceutical therapies and secondary for research purposes. Although RCTs may have not contributed enough in the improvement of mental health indices or overall costs of treatment, they still represent the best research practice available.

Yet, it is noteworthy that a gap exists between research and clinical practice. Patients who participate in clinical trials, especially placebo controlled trials, may not be representative of "real life patients", considering that more severe patients (e.g. with suicidal ideation) are usually excluded. Also, participants are likely to differ in different study sites- e.g. EU and USA patients may be different, which also determines different placebo group response rates, as seen for example in Brexpiprazole study (Vieta et al., 2021). Also, in developing countries, clinical trials may be the way for patients to receive treatments which wouldn't be available for them otherwise, which could account for recruitment of more severe patients, and this also may impact the generalizability of the results. In addition, the relatively short duration of RCTs makes it difficult to formulate a solid decision for clinical practice on a new drug based on RCTs only, and although they are a first step necessary for drug approval. Furthermore, most studies are funded by pharmaceutical companies and drug research by the industry can be different from drug research by the academic centers (Lundh et al.,2017). Yet, with all these limitations, RCTs represent the only option for regulatory and research purposes of any new drug.

The present study demonstrates the extent of the consensus among a large number of international experts. It provides timely objective data on a topic that is often disputed on the basis of emotions and private beliefs and not necessarily on evidence. Its limitations include that consensus is not an absolute measure and can vary with scientific advances and new data. Delphi studies are meant to be subjective, so that they can reflect the status of current clinical opinion. Between the two rounds of rating following the Delphi protocol, 8/41 participants did not respond to the second survey (Fig. 1), which could have produced selection bias and influenced the results for the re-rated items. The choice of experts and issues such as the wording of statements might have influenced the results. Most of the experts included are clinical researchers, and most are psychiatrists with clinical experience, which may be reflected in the recommendations.

In summary, the endorsed recommendations established by our Delphi study can inform trial-design choice in psychopharmacological trials and beyond. The most salient of these recommendations includes not excluding a priori the use of placebo-controlled trials. The expert panel agreed that this experimental design is the most reliable and still facilitates the introduction of new pharmaceuticals to the market for improving patient wellbeing. Regarding active-controlled trials, there was consensus that they play an integral role in approving drugs with equal efficacy but different tolerability. Non-inferiority trials were the least recommended trial design. Experts also agreed that trial-design choice must balance the risk of approving ineffective medications with the risk of rejecting valuable new treatments. The present recommendations represent consensus of 41 experts regarding the use of different trial designs for drug approval trials in psychiatry, and strongly support the use of randomized clinical trials in general, and the use of placebo-controlled trials in particular.

Role of funding source

No funding was provided for this study.

Contributors

MS and CP contributed to the literature searches and preparation of the manuscript. FK contributed to the literature search, Delphi study, data collection and data analysis. DP prepared the study protocol, contributed to the data collection, analysis, writing and editing of all manuscript drafts. All remaining authors contributed to the writing and editing of the manuscript. All authors contributed to and approved the final manuscript.

Declaration of Competing Interest

Guy M. Goodwin is a NIHR Emeritus Senior Investigator, holds shares in P1vital and P1Vital products and has served as consultant, advisor or CME speaker in the last 3 years for Beckley Psytech, Clerkenwell Health, Compass pathways, Evapharma, Janssen, Lundbeck, Medscape, Novartis, P1Vital, Sage, Servier. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. Michael Berk is supported by a NHMRC Senior Principal Research Fellowship (1156072). MB has received Grant/Research Support from the NIH, Cooperative Research Centre, Simons Autism Foundation, Cancer Council of Victoria, Stanley Medical Research Foundation, Medical Benefits Fund, National Health and Medical Research Council, Medical Research Futures Fund, Beyond Blue, Rotary Health, A2 milk company, Meat and Livestock Board, Woolworths, Avant and the Harry Windsor Foundation, has been a speaker for Abbot, Astra Zeneca, Janssen and Janssen, Lundbeck and Merck and served as a consultant to Allergan, Astra Zeneca, Bioadvantex, Bionomics, Collaborative Medicinal Development, Janssen and Janssen, Lundbeck Merck, Pfizer and Servier – all unrelated to this work. Andreas Meyer-Lindenberg has received consultant fees from: Boehringer Ingelheim, Elsevier, Brainsway, Lundbeck Int. Neuroscience Foundation, Lundbeck A/S, The Wolfson Foundation, Bloomfield Holding Ltd, Shanghai Research Center for Brain Science, Thieme Verlag, Sage Therapeutics, v Behring Röntgen Stiftung, Fondation FondaMental, Janssen-Cilag GmbH, MedinCell, Brain Mind Institute, Agence Nationale de la Recherche, CISSN (Catania Internat. Summer School of Neuroscience), Daimler und Benz Stiftung, American Association for the Advancement of Science, Servier International. Additionally, he has received speaker fees from: Italian Society of Biological Psychiatry, Merz-Stiftung, Forum Werkstatt Karlsruhe, Lundbeck SAS France, BAG Psychiatrie Oberbayern, Klinik für Psychiatrie und Psychotherapie Ingolstadt, med Update GmbH, Society of Biological Psychiatry, Siemens Healthineers, Biotest AG -All unrelated to this work. Terence A. Ketter has been a consultant to Otsuka Pharmaceuticals, Sunovion Pharmaceuticals, Abbvie, and Alkermes. Joseph Goldberg has been a consultant to BioXCel, Otsuka, Sage Pharmaceuticals, Sunovion, and WebMD, and served on the speaker boards for Allergan, Intracellular Therapies, and Sunovion. Rif S. El-Mallakh is on the speaker bureau of Alkermes, Eisai, Indivior, Intra-Cellular Therapeutics, Janssen, Lundbeck, Noven, Otsuka, Sunonvion, and Teva. Lakshmi N Yatham has been on speaker/advisory boards for, or has received research grants from Abbvie, Alkermes, Allergan, CANMAT, CIHR, DSP, Merck, and Sanofi Rasmus W. Licht has within the preceding three years served an advisory board of Janssen Cilag and Sagw, and received speaker honorarium from Astra-Zeneca, Jannsen-Cilag, Servier and Lundbeck Allan H. Young has been employed by King's College London; Honorary Consultant SLaM (NHS UK). Young has participated in paid lectures and advisory boards for the following companies with drugs used in affective and related disorders: Astrazenaca, Eli Lilly, Lundbeck, Sunovion, Servier, Livanova, Janssen, Allegan, Bionomics, Sumitomo Dainippon Pharma, COMPASS Allan H. Young's independent research funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health.Consultant to Johnson & Johnson and Livanova. Received honoraria for attending advisory boards and presenting talks at meetings organized by LivaNova. Prof. Alessandro Serretti is or has been consultant/speaker for: Abbott, Abbvie, Angelini, Astra Zeneca, Clinical Data, Bo- heringer, Bristol Myers Squibb, Eli Lilly, GlaxoSmithKline, Innovapharma, Italfarmaco, Janssen, Lundbeck, Naurex, Pfizer, Polifarma, Sanofi, Servier. Ayşegül Yildiz has nothing to declare. Jean Michel Azorin has received honoraria or research or educational conference grants from Lundbeck and Otsuka. Othman Sentissi has received advisory board honoraria or research or educational conference from Otsuka, Lilly, Lundbeck, Sandoz, Janssen and Sunovion on an institutional account for research and teaching. Prof. Gordon Parker has received lecture fees and board member honoraria from Otsuka, Servier and Lundbeck. Dr. David Bond has received consulting fees and/or research grants from Alkermes, Myriad Genetics, the National Institutes of Health, the University of Minnesota Department of Psychiatry and Behavioral Sciences, and the University of Minnesota Foundation. Prof. Giulio Perugi has received grant/research support from Eli Lilly & Co.; is on the speaker/advisory board of Sanofi-Aventis, Bristol-Myers Squibb, AstraZeneca, Eli Lilly & Co., Jannsen-Cilag, and Lundbeck; and has acted as consultant of AstraZeneca, Eli Lilly & Co., and Lundbeck. Prof. Eduard Vieta has received grants and served as consultant, advisor or CME speaker for the following entities: AB- Biotics, Abbott, Allergan, Angelini, Dainippon Sumitomo Pharma, Galenica, Janssen, Lundbeck, Novartis, Otsuka, Sage, Sanofi-Aventis, and Takeda. Dina Popovic has served as a speaker and/or medical writer and/or consultant and/or has participated in advisory boards for Bristol-Myers Squibb, Dexel, Merck Sharp & Dohme, Janssen-Cilag, Lundbeck, Ferrer, and Forum Pharmaceuticals. None of the remaining authors have conflicts of interest to declare.

Acknowledgments

We do not have acknowledgments.

References

Barbui, C., Bighelli, I., 2013. A new approach to psychiatric drug approval in Europe. PLoS Med. 10 (10).

Batra, S., Howick, J., 2017. Empirical evidence against placebo controls. J. Med. Ethics 43 (10), 707–713.

Burger, H.U., Beyer, U., Abt, M., 2011. Issues in the assessment of non-inferiority: perspectives drawn from case studies. Pharm. Stat. 10 (5), 433–439.

D'Agostino, R.B., Massaro, J.M., Sullivan, L.M., 2003. Non-inferiority trials: design concepts and issues - the encounters of academic consultants in statistics. Stat. Med. 22 (2), 169–186.

Dalkey, N., Helmer, O., 1963. An experimental application of the Delphi method to the use of experts. Manag. Sci. doi:10.1287/mnsc.9.3.458.

de Craen, A.J.M., Kaptchuk, T.J., Tijssen, J.G.P., Kleijnen, J., 1999. Placebos and placebo effects in medicine: Historical overview. J. Roy. Soc. Med. 92 (10), 511–515.

Ellenberg, S.S., Temple, R., 2000. Placebo-controlled trials and active-control trials in the evaluation of new treatments: Part 2: Practical issues and specific cases. Ann. Intern. Med. 133 (6), 464.

Fleischhacker, W.W., Czobor, P., Hummer, M., Kemmler, G., Kohnen, R., Volavka, J., 2003. Placebo or active control trials of antipsychotic drugs? Arch. Gen. Psychiatry 60 (5), 458–464.

Garattini, S., Bertele, V., 2009. Ethics in clinical research. J. Hepatol. 51 (4), 792–797.

Glass, K.C., 2008. Ethical obligations and the use of placebo controls. Can. J. Psychiatry 53 (7), 428–429.

Hasnain, M., Rudnick, A., Bonnell, W.S., Remington, G., Lam, R.W., 2018. Use of placebo in clinical trials of psychotropic medication. Can. J. Psychiatry 63 (5), 338–341.

Healy, D., 2003. The Creation of Psychopharmacology. Schweizer Archiv Für Neurologie Und Psychiatrie 154 (01), 43.

Howick, J., 2009. Questioning the methodologic superiority of "Placebo" over "Active" controlled trials. Am. J. Bioethic. 9 (9), 34–48.

Lundh, A, Lexchin, J, Mintzes, B, Schroll, JB, Bero, L., 2017 Feb 16. Industry sponsorship and research outcome. Cochrane Datab. Syst. Rev. 2 (2), MR000033.

Kemp, A.S., Schooler, N.R., Kalali, A.H., Alphs, L., Anand, R., Awad, G., Davidson, M., Dubé, S., Ereshefsky, L., Gharabawi, G., Leon, A.C., Lepine, J.P., Potkin, S.G., Vermeulen, A., 2010. What is causing the reduced drug-placebo difference in recent schizophrenia clinical trials and what can be done about it? Schizophr. Bull. 36 (3), 504–509.

Khan, A., Fahl Mar, K., Faucett, J., Khan Schilling, S., Brown, W.A, 2017. Has the rising placebo response impacted antidepressant clinical trial outcome? Data from the US Food and Drug Administration 1987-2013. World Psychiatry 16 (2), 181–192.

Krol, F.J., Hagin, M., Vieta, E., Harazi, R., Lotan, A., Strous, R.D., Lerer, B., Popovic, D., 2020. Placebo—To be or not to be? Are there really alternatives to placebo-controlled trials? Eur. Neuropsychopharmacol. 32, 1–11.

Levine, R.J., 1999. The need to revise the declaration of Helsinki. N. Engl. J. Med. 341 (7), 531–534.

Linstone, H.A., Turoff, M., 1975. The Delphi method: techniques and applications. J. Market. Res. 18 (3), 317.

Lintonen, T., Konu, A., Rönkä, S., Kotovirta, E., 2014. Drugs foresight 2020: a Delphi expert panel study. Subst. Abuse 9 (1).

Murawiec, S., Popovic, D., 2015. The ultimate need of "precision psychiatry" Interview with Dr Dina Popovic. Psychiatria 12 (4), 281–282.

Nolen, W.A., Licht, R.W., Young, A.H., Malhi, G.S., Tohen, M., Vieta, E., Kupka, R.W., Zarate, C., Nielsen, R.E., Baldessarini, R.J., Severus, E., 2019. What is the optimal serum level for lithium in the maintenance treatment of bipolar disorder? A systematic review and recommendations from the ISBD/IGSLI Task Force on treatment with lithium. Bipolar Disord. 21 (5), 394–409.

Ovosi, J.O., Ibrahim, M.S., Bello-Ovosi, B.O., 2017. Randomized controlled trials: Ethical and scientific issues in the choice of placebo or active control. Ann. Afr. Med. 16 (3), 97–100.

Pacchiarotti, I., Bond, D.J., Baldessarini, R.J., Nolen, W.A., Grunze, H., Licht, R.W., Post, R.M., Berk, M., Goodwin, G.M., Sachs, G.S., Tondo, L., Findling, R.L., Youngstrom, E.A., Tohen, M., Undurraga, J., González-Pinto, A., Goldberg, J.F., Yildiz, A., Altshuler, L.L., … Vieta, E., 2013. The International Society for bipolar Disorders (ISBD) task force report on antidepressant use in bipolar disorders. Am. J. Psychiatry 170 (110), 1249–1262.

Pocock, S.J., 2003. The pros and cons of noninferiority trials. Fundam. Clin. Pharmacol. 17 (4), 483–490.

Pocock, S.J., Stone, G.W., 2016. The primary outcome fails — what next? N. Engl. J. Med. 375, 861–870.

Popovic, D., Benabarre, A., Crespo, J.M., Goikolea, J.M., González-Pinto, A., Gutiérrez-Rojas, L., Montes, J.M., Vieta, E., 2014. Risk factors for suicide in schizophrenia: Systematic review and clinical recommendations. Acta Psychiatr. Scand. 130 (6), 418–426.

Popovic, D., Reinares, M., Goikolea, J.M., Bonnin, C.M., Gonzalez-Pinto, A., Vieta, E., 2012. Polarity index of pharmacological agents used for maintenance treatment of bipolar disorder. Eur. Neuropsychopharmacol. 22 (5), 339–346.

Popovic, D., Vieta, E., Azorin, J.M., Angst, J., Bowden, C.L., Mosolov, S., Young, A.H., Perugi, G., 2015. Suicide attempts in major depressive episode: Evidence from the BRIDGE-II-Mix study. Bipolar Disord. 17 (1), 795–803.

Powers, J.H., Cooper, C.K., Lin, D., Ross, D.B., 2005. Sample size and the ethics of non-inferiority trials. The Lancet 366 (9479), 24–25.

Shapiro, S., Fergusson, D., Cranley Glass, K., 2010. Substituting placebo for established, effective therapy: Why not? CMAJ 182 (16), 1749–1753.

Snapinn, S.M., 2014. Noninferiority Trial. Curr. Control Trial. Cardiovasc. Med. 1 (1), 19–21.

Spielmans, G.I., Kirsch, I., 2014. Drug approval and drug effectiveness. Ann. Rev. Clin. Psychol. 10 (1), 741–766.

Spławiński, J., Kuźniar, J., 2004. Clinical trials: active control vs placebo - what is ethical? Sci. Eng. Ethics 10 (1), 73–79.

Streiner, D.L., 1995. The ethics of placebo-controlled trials. Can. J. Psychiatry 40 (4), 165–166.

Streiner, D.L., 2008. The lesser of 2 evils: The ethics of placebo-controlled trials. Can. J. Psychiatry 53 (7), 430–432.

Temple, R., Ellenberg, S.S., 2000. Placebo-controlled trials and active-control trials in the evaluation of new treatments: Part 1: Ethical and scientific issues. Ann. Intern. Med. 133 (6), 455–463.

Vieta, E., Cruz, N., 2012. Head to head comparisons as an alternative to placebo-controlled trials. Eur. Neuropsychopharmacol. 22 (11), 800–803.

Vieta, E, Sachs, G, Chang, D, Hellsten, J, Brewer, C, Peters-Strickland, T, Hefting, N, 2021 Aug. Two randomized, double-blind, placebo-controlled trials and one open-label, long-term trial of brexpiprazole for the acute treatment of bipolar mania. J. Psychopharmacol. 35 (8), 971–982.

Walsh, T.B., Seidman, S.N., Sysko, R., Gould, M., 2002. Placebo response in studies of major depression: variable, substantial, and growing. J. Am. Med. Assoc. 287 (14), 1840–1847.

World Medical Association, 1996. WMA Declaration of Helsinki—Ethical Principles for Medical Research Involving Human Subjects.

World Medical Association, 2018. WMA Declaration of Helsinki—Ethical Principles for Medical Research Involving Human Subjects.