



**HAL**  
open science

## **Project Rebuild the Evidence Base (REB): a method to interpret randomised clinical trials and their meta-analysis to present solid benefit-risk assessments to patients**

Rémy Bousageon, Clara Blanchard, Elodie Charuel, Thibault Menini, Bruno Pereira, Florian Naudet, Behrouz Kassai, François Gueyffier, Michel Cucherat, Hélène Vaillant-Roussel

### ► **To cite this version:**

Rémy Bousageon, Clara Blanchard, Elodie Charuel, Thibault Menini, Bruno Pereira, et al.. Project Rebuild the Evidence Base (REB): a method to interpret randomised clinical trials and their meta-analysis to present solid benefit-risk assessments to patients. *Therapies*, 2023, 78 (4), pp.353-365. 10.1016/j.therap.2022.10.008 . hal-03829199

**HAL Id: hal-03829199**

**<https://uca.hal.science/hal-03829199v1>**

Submitted on 3 Feb 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

**THERAPIES****CLINICAL PHARMACOLOGY****Project Rebuild the Evidence Base (REB): a method to interpret randomised clinical trials and their meta-analysis to present solid benefit-risk assessments to patients**

Project REB: rebuilding the “E” in EBM

**Rémy Bousageon** <sup>a,b,c</sup>, **Clara Blanchard** <sup>d,\*</sup>, **Elodie Charuel** <sup>e</sup>, **Thibault Menini** <sup>e</sup>, **Bruno Pereira** <sup>f</sup>, **Florian Naudet** <sup>g</sup>, **Behrouz Kassai** <sup>b</sup>, **François Gueyffier** <sup>b</sup>, **Michel Cucherat** <sup>b</sup>, **Hélène Vaillant-Roussel** <sup>e</sup>

<sup>a</sup> *Collège universitaire de médecine générale, UCBL, Lyon 1, 69000 Lyon, France*

<sup>b</sup> *UMR 5558, LBBE - EMET, CNRS – UCBL, Lyon 1, 69000 Lyon, France*

<sup>c</sup> *Conseil scientifique du CNGE, 75000 Paris, France*

<sup>d</sup> *Department of General Practice, Poitiers University, 86073 Poitiers, France*

<sup>e</sup> *Department of General Practice and Research Unit ACCePPT, Clermont Auvergne University,*

*63001 Clermont-Ferrand, France*

<sup>f</sup> *CHU Clermont-Ferrand, Biostatistics Unit (Department of Clinical research and Innovation),*

*63000 Clermont-Ferrand, France*

<sup>g</sup> *Univ Rennes, CHU Rennes, Inserm, CIC 1414 [(centre d’investigation clinique de Rennes)],*

*35000 Rennes, France*

Received 27 December 2021; accepted 3 October 2022

**\*Corresponding author.** Department of General Practice, Poitiers University, 15 rue de l'Hôtel Dieu - TSA 71117, 86073 Poitiers cedex 9, France.

*E-mail address:* Cla.blanchard@gmail.com (C. Blanchard)

## Summary

Evidence-based medicine is the cornerstone of shared-decision making in healthcare today. The public deserves clear, transparent and trust-worthy information on drug efficacy. Yet today, many drugs are prescribed and used without solid evidence of efficacy. Clinical trials and randomized clinical trials (RCTs) are the best method to evaluate drug efficacy and side effects. In a shared medical decision-making approach, general practitioners need drug assessment to be based on patient-important outcomes. The aim of project Rebuild the Evidence Base (REB) is to bridge the gap between the data needed in clinical practice and the data available from clinical research. The drugs will be assessed on clinical patient important outcomes and for a population. Using the Cochrane tools, we propose to analyse for each population and outcome : 1) a meta-analysis based on RCTs with a low risk of bias overall ; 2) an evaluate of results of confirmatory RCTs; 3) a statistical analysis of heterogeneity between RCTs, and 4) an analysis of publication bias. Depending on the results of these analyses, the evidence will be categorised in 4 different levels: firm evidence, evidence (to be confirmed), signal or absence of evidence. Project REB proposes a method for reading and interpreting randomized clinical trials and their meta-analysis to produce quality data for general practitioners to focus on benefit-risk assessment in the interest of patients. If this data does not exist, it could enable clinical research to better its aim.

## KEYWORDS

Health-policy; Evidence-based medicine; Meta-analysis; Patient safety; Communication

## Abbreviations

RCT: randomized clinical trials

REB: Rebuild the Evidence Base

## Introduction

For the last 30 years, evidence based medicine (EBM) has upheld that scientific data and its level of certainty should be the cornerstone of decision-making in everyday practice [1]. Although it has allowed undeniable progress in the fields of research and medicine, this approach is being challenged [2-6].

Firstly, many authors [4,6] lament the diversion of EBM away from the interest of patients. Indeed, it is uncertain whether some research data [4] and guidelines are trust-worthy [7], particularly when financial gains are in play. Results may be manipulated by pharmaceutical firms [8-20], when experts have conflicts of interest [21-25] or when there is publication [26,27], or cognitive or interpretation biases [28,29] (for example “optimism bias” [30] or confirmation bias i.e “seeing what you want see” [31] bias). Spins in interpretation and presentation of results can cloud the results, even of studies with a supposed « high level of evidence» such as meta-analyses of randomised controlled trials (RCTs) [4,6,32,33]. For example, in a sample of 93 RCTs in the cardio-vascular field, an estimated 54% (95% CI : 44%-64%) [34] had a spin in the conclusions (a positive interpretation of statistically insignificant results).

Secondly, although proponents of EBM have highlighted the importance of finding the « best available evidence » [1,35] by using RCTs and their meta-analyses rather than observational studies [36,37] for decision-making, they have not defined what makes data sufficiently convincing to justify, for example a drug prescription. This has led to a great variability in the interpretation of what is evidence and so in decisions made in every day practice and recommendations from guidelines [38] (Box 1). Even if RCTs or meta-analyses of RCTs are the best evidence available, what is the methodological minimum required to consider the evidence to be compelling enough to propose a therapy to a patient? Could firm evidence derive from a high risk of bias clinical trial or a meta-analyses of studies low-quality results? Can significant results with a “*very low quality of evidence*” [39] be considered convincing? Haynes et al. consider results of low quality RCTs on par with those of observational studies [40]. Djulbegovic and Guyatt recently wrote « *prudence generally dictates not implementing interventions when only very low quality evidence exists* » [41]. However, only up to 8 to 12% of therapies have proven their efficacy with certainty [42-44] for patient-important outcomes (PIO). For example, among 201 systematic reviews of pharmacological intervention, only 12.4% had a positive conclusion for a PIO with a high level of evidence [42].

Another pivotal point is the trust a health-professional has in available drugs. If doctors prescribe a drug to their patients, it is because they believe it has been authorised by a higher authority (like the Food and Drug Administration [FDA] or the European Medicines Agency [EMA]) and so that its therapeutic benefit must be established, at least for the indication approved in the guidelines. In real life, doctors have neither the time[45], nor the skill [46] to correctly evaluate the evidence and so put their trust in national policies or guidelines [47]. In practice, decisions are often not based on solid evidence (Box 2 for examples). How can we promote a treatment with no proven benefit within the “*primum non nocere*” ethical principle?

Before asking what the “best **available** evidence” is, should we not decide what is the “best **possible and necessary** evidence”? What is the best possible and strict minimum needed to justify a prescription and so what does the E in EBM really represent ? Should we not hurry to answer this question before the use of unconvincing evidence in decision-making and the hacking of EBM has lasting repercussions on the relationship between

health-care givers and patients? How can patients continue to trust in the scientific method and health institutions and therefore in the doctors holding the prescribing pens?

In our opinion, the method behind EBM remains the best available for health-care providers and their patients. In fact, many authors have underscored its crucial role regardless of its limitations and have called for the rebuilding of EBM on more solid bases [6,48]. Indeed, it allows a rational use of care for example by limiting over-treatment [49] and allowing a justified halt to unvalidated common practices when RCTs do not prove the benefit of a drug [50]. But there can be no truly informed decision without deciding whether or not the evidence is conclusive for a specific drug in a specific context [51,52]. For the moment, to our knowledge, there is no tool or database that allows general practitioners and patients to know what they can really expect of a drug in a specific clinical situation expressed in terms of effect-size on a PIO.

In keeping with the call to rebuild the evidence-base [2,6,48], and because doctors cannot evaluate the soundness of multiple guidelines [53-55] or of multiple meta-analyses on the same topic [56,57], we aim to propose a solid and objective method for reading and interpreting data (from RCTs or meta-analyses of RCTs) about the treatment effect of drugs. We call this the REB method: Rebuild the Evidence-Base [58] in line with the EBM manifesto [6] We hope it will help restore trust in drugs, in medicine and in doctors by providing the right information for shared-decision making.

## **Methods**

### **Protocol**

The protocols will be registered, for each drug of interest, in a secure database such as PROSPERO (<https://www.crd.york.ac.uk/PROSPERO/>) or other (eg. <https://osf.io/registries>) and will follow the P-Prisma guidelines [59]. The procedure for these reviews will follow established best methods and Cochrane standards [60].

Drugs chosen initially will be specific molecules used in every day practice in general practice, notably drugs prescribed for comfort, such as for pain control or benign symptoms. But the method could be used for any type of drug, no matter the indication.

### **Literature search**

We will search at least in MEDLINE, Cochrane Central Register of Controlled Trials (CENTRAL) and Embase without any language restriction, but others database (Scopus, Web of Science, etc ) may be used. We will combine both MeSH and free text terms to identify relevant articles (example international non-proprietary name). Once the MEDLINE strategy is determined, it will be adapted to the syntax and subject headings of the other

databases. We will also search the publication of study protocols : at least in ClinicalTrials.gov, but others also such as the World Health Organization (WHO) International Clinical Trials Registry Platform (ICTRP) Search Portal (<https://trialsearch.who.int/>) to detect ongoing or unpublished trials in order to assess publication bias. And, when possible, Cochrane's systematic review on the same topic will be consulted so as to identify additional eligible studies. If strong reviews of literature have already been published on the same population, drug and PIO, we will update these reviews and the entire process described below will be applied to the previously included and newly included studies.

### **Eligibility criteria**

We will include RCTs testing for superiority of the drug of interest against placebo or active drugs in a specific indication and/or population. The outcomes will be PIO (not surrogate). Non-inferiority trials testing against the standard drug may also be considered. In this case, we will check for clinical pertinence by evaluating the potential loss of efficacy in balance with the benefits of the drug in question.

### **Study process**

Teams of two paired reviewers, without any conflicts of interest (<http://www.transparence.gov>) trained in health research methods, will independently screen titles, abstracts and full texts for eligibility. They will assess risk of bias and collect data from each eligible study, using standardised, pilot tested forms, together with detailed instructions. Reviewers will resolve disagreement through discussion or, if required, by arbitration by a third reviewer.

### **Data collection**

We will collect the following information from each eligible RCT:

- General study characteristics: author name, year of publication, total number of patients randomised, number of treatment groups, length of follow-up, study phase, funding source, trial registry number, countries involved and number of study sites;
- Patient characteristics: gender, age and pertinent characteristics according to the drug of interest and the target population;
- Intervention: baseline treatment, details of intervention and control groups (eg, International Non-proprietary Name of drug, dosage and duration of treatment) and concomitant treatments throughout the study;
- Comparator: placebo or active drug (standard or not);
- Outcomes: how primary and secondary outcomes are defined and measured in each study or protocol, number of events and patients included for analyses in each group, as well as adjusted data if available.

### **Taking into account the risk of false positives**

If possible, the initial protocol (published or not) of the RCT will be used to better evaluate the study's biases, particularly bias due to deviations from the intended interventions or statistical analysis plan. This will verify that the hypothetico-deductive approach is respected and that the result of the RCT is confirmatory, rather than exploratory [61-63].

### **Risk of bias assessment**

We will assess the risk of bias of included studies with the Cochrane Collaboration's ROB's V2 tool [64] bias arising from the randomisation process, bias due to deviations from intended interventions, bias due to missing outcome data, bias in measurement of the outcome, bias in selection of the reported result. These will be judged as low, some concern or high risk for each type of bias for a specific outcome.

### **Risk associated to multiplicity of statistical tests and type I error**

In case of multiple testing of the main outcome (for example in case of multiple outcomes or intermediary analyses), we will check for adjustment methods [61,65-67] to control the risk of a type I error (examples : Bonferroni's correction or outcome that were planned in hierarchal analysis). If these analyses are done we consider the risk of a type I error is controlled. If this risk is not anticipated in the protocol, the risk of false positivity is high and the result of the study will then be considered as exploratory, even if a statistical difference is found [67].

If the primary outcome is a composite criterion, we will consider statistically significant results of an individual outcome of this composite criterion as a primary outcome with a controlled risk of type I error after checking for the risk of multiple testing (for example if planned in hierarchical analysis) or after checking for consistency.

### **Definition of a "conclusive" RCT = confirmatory trial and overall low risk of bias**

We will consider that an RCT is conclusive for the outcomes chosen if these two criteria are true:

- The **overall risk of bias is low**, and

- The results are statistically significant **for outcomes** that were defined a priori (or that were modified without any knowledge of the results) and after taking into account the risk of multiple testing and type I error (in confirmatory trials [63]).

### **meta-analysis**

For each outcome, a meta-analysis will take into consideration only results from RCTs with an overall low risk of bias. The result of this meta-analysis will be considered positive if the results are statistically significant ( $p < 0.05$ ).

### **statistical analysis**

For all outcomes, the number of events in each group will be extracted. As much as possible, we will pool the risk ratios reported in the studies as these are often hazard ratios or adjusted risk ratios. For dichotomous (binary or qualitative) outcomes, we will calculate Risk Ratios (RR) with 95% CIs from the number of events and participants in each treatment group. For continuous (quantitative) outcomes, we will calculate the difference in means (or standardised mean differences if necessary) with 95% confidence intervals (CIs). If events are very rare (0 or 1 in each group of the RCT), we will pool data using Peto's methods (because corrections for zero cell counts are not necessary) with reported pooled Peto odds ratios and associated 95% CIs. We will use a random-effect model meta-analysis to assess the treatment-effect of a drug. Statistical heterogeneity across trials will be assessed with the  $I^2$  statistic. The  $I^2$  statistic measures the proportion of overall variation that is attributable to between-study heterogeneity and according to the Cochrane Handbook [60] the uncertainty in the value of  $I^2$  is important when the number of studies is low.

Heterogeneity is generally considered substantial if the  $I^2$  is between 50% and 90% (and considerable between 75% and 100%) but the interpretation of heterogeneity and its derogatory character will be left to the judgement of the authors. When possible, we will use the results from intention-to-treat (ITT) analyses from the RCTs included. Publication bias will be evaluated directly by analysing trial registries and when the number of trials is sufficient by Egger's Test.

### **Quality of evidence**

In the end, the number of conclusive RCTs, the heterogeneity and the publication bias between the RCTs will be evaluated. Publication bias and heterogeneity will be evaluated' each time but the analysis of the results (is it a significant source of bias or not?) will be left to the authors' discretion. The process and a guide to interpretation are presented Fig. 1.



**For each population and each outcome**, the REB method will conclude to one of four levels:

**1. Firm evidence (result with little chance of being refuted):**

At least 2 conclusive RCTs for the outcome chosen;

AND No suspected or real publication bias;

AND positive meta-analysis for the outcome chosen without substantial heterogeneity.

**2. Evidence: data is convincing and must be confirmed so as to reach “firm evidence” level:**

Only 1 conclusive RCTs for the outcome chosen;

AND No suspected or real publication bias;

AND positive meta-analysis for the outcome chosen without substantial heterogeneity.

**3. Signal: results to be confirmed but no evidence:**

Results are promising and need to be confirmed by a conclusive RCT focusing on the same outcome.

A) No conclusive RCT for the outcome chosen;

AND positive meta-analysis for the outcome chosen without substantial heterogeneity in the meta-analysis (if meta-analysis possible);

AND No suspected or real publication bias.

B) At least 1 conclusive RCTs for the outcome chosen AND one of these situations:

- Not positive meta-analysis (negative or inconclusive);
- OR positive meta-analysis but substantial heterogeneity in the meta-analysis;
- OR positive meta-analysis but suspected or real publication bias.

**4. Lack of evidence**

If an analysis does not fit one of the above cases, it will be considered that there is not enough evidence to conclude on the effect of the drug.

Here are some cases where this conclusion will be applied:

A) No RCTs included after data search.

B) No conclusive RCT and one of the following:

- Not positive meta-analysis (if possible);
- OR positive meta-analysis but substantial heterogeneity in the meta-analysis;
- OR positive meta-analysis but suspected or real publication bias;

C) At least 1 conclusive RCTs for the outcome chosen AND one of this situation:

- Not positive meta-analysis with substantial heterogeneity OR suspected or real publication bias;

- Positive meta-analysis but substantial heterogeneity in the meta-analysis AND suspected or real publication bias.

If the meta-analysis cannot be carried out, the decision will fall to the authors as it will depend on the situations and the trials available, their statistical strength and the data available.

## Discussion

Through the application of a pre-determined process, REB aims to deliver the truest interpretation of drug efficacy. The highest level of evidence, firm evidence, will prove a specific causal effect of the drug, with a very low probability of rebuttal [68]. This process is adapted to discuss drug efficacy, and not toxicity for which other types of studies are warranted (such as observational studies [69]). Indeed, the reasoning is different when discussing risk evaluation as “signals” must be examined along with their overall coherence. In the same way, the REB method is probably not adapted to all treatments.

### **REB: why is it important and what is its position in relation to other methods?**

Firstly, in contrast with other methods [70-72], with GRADE [37,68,73], or with trial sequential analysis [72,74], REB highlights the fundamental role of confirmatory RCTs [61-63] in deciding whether the data is convincing or not, especially in drug therapies (Box 3). Indeed, this type of study is the most appropriate to discuss efficacy and decision-making. Hence why these studies are the basis of most regulation agencies’ decisions where dichotomy is necessary (to prescribe or not to prescribe, to approve or not to approve). REB is a standardised way to present and interpret results to doctors and patients. Hypothesis generating results from RCTs or meta-analyses are not considered as evidence, in this way REB differs from the classic meta-analytical approach. Convincing data must come from statistically significant results of a positive meta-analysis with low heterogeneity and no publication bias and with at least a confirmatory RCT conclusive for a PIO (with control of type I error), with a low risk of bias and in line with other RCTs. Strictly speaking, firm evidence must come from two confirmatory RCTs (Box 4) along with the above-mentioned criteria in agreement with the scientific principle of reproducibility [70]. The REB method focuses on the body of evidence [73] to allow a truer interpretation of scientific results.

Secondly, REB focuses on RCTs of high quality (low overall risk of bias). This is essential to lower the risk of false positives and so the treatment of patients with useless, costly or even harmful drugs. The process highlights the importance of risk of bias evaluation although rating the quality of evidence is partly subjective [75]. However, ROB2 describes the evaluation criteria explicitly and transparently so as to harmonise evaluations and minimise the arbitrary of judgments [64]. Indeed, it is worth noting that although the importance of taking into account risk of bias evaluation is well established [60,64], in practice Hopewell et al. found that sensitivity analyses according to risk of bias is only done in about 7% of systematic reviews

of RCTs [76], which supports the fact that the risk of bias of studies is not often taken into account (Box 1). REB proposes to reverse the usual reasoning so as to analyse only the unbiased results of high quality RCTs. Trials with unacceptable risk of bias are to be considered as exploratory, and results cannot be convincing.

Finally REB will determine the effect of a drug on a specific PIO and in a specific population. Indeed, the “evidence” in EBM must be centered on the patient to sway medical decisions. It underlines that proof is only convincing in a specific context depending on the PICO (population, intervention, control, outcome) characteristics [77] and extrapolation can never replace direct proof [78-80]. This is essential as a recent study showed that as many as 36% of subgroup characteristics interacted with the treatment effect. Also a drug that may have proven its efficacy in one context, may be less efficient in a different one. This is true for example for pregabalin which is considered efficient for neuropathic pain [39] but not for chronic back pain or sciatalgia [81,82]. Similarly antidepressants have yet to prove their efficacy in specific contexts like alcoholism [83], cancer [84] or Alzheimer's disease [85,86]. Contrarily to I-DPP4 [87-88], GLP1-agonists seem to lower the cardiovascular risk of patients [89]. Even if their mechanisms seem to put them in the same class (incretin pathway), it is not justified to extrapolate that I-DPP4 could have the same benefits [90]. In this way, applying a result to a patient is a leap of faith [77] because the patient cannot be identical to that included in RCTs. In our opinion, this leap of faith can be part of shared decision-making but not of a scientific demonstration.

### **Significance of reb in practice: informing doctors and patients of the reality of the evidence and of sufficiently convincing facts**

REB answers two questions “Can it work?” And “Does it work in practice?” [91] and these are the first stepping stones essential to a truly shared-decision. These first steps are necessary but not enough as the level of evidence cannot suffice to decide to treat or not a patient [92-94]. Other factors intervene such as the clinical pertinence (effect size), the harm, the burden, the cost or the availability of alternatives. After these drug-centered considerations, there is also the patient's beliefs and values and the posture and experience of the doctor [92]. REB is a method for reading and interpreting data, from RCTs and meta-analyses, to define how firm the evidence is, but the leap from evidence to decision depends on the quality of the patient-doctor relationship. Indeed, science says “what is” and not “what should be done”. If the decision can vary according to the context (depending on the patient's choice or the prognosis or the available alternatives...), the solidity of the evidence does not. With a logical, explicit and scientific approach, we can say if a result is evidence or not. Choosing the appropriate method to answer a research question and deciding a priori how to interpret the results can generate convincing data. Without this method, the interpretation is much more influenced by preconception, context and other biases (notably financial ones).

Seeing as high-quality evidence in favour of clear therapeutic benefits is rare, we do not expect to find a lot of “firm evidence” or “evidence” of drug efficacy, especially as the criteria is very demanding [42-44]. Although there may be few drugs that meet this requirement, this is not a good reason not to do this demanding assessment. We consider it of utmost importance that time and money should not be wasted on RCTs that are not designed to minimise the risk of bias as much as possible. Medical research owes it to the public to provide information on drug efficacy that is as transparent and irrefutable as possible. By lifting the fog [93], the results of REB could help doctors, patients, teachers, regulators and public health decision-makers. In practice and for shared decision-making, it can be legitimate to treat a patient even though there isn't a convincing level of evidence of efficacy, for example in case of pain or when no better alternative is available and when no important side effects exist [94-95]. In the

same way, the extrapolation and the applicability of a proven effect on a certain PIO can be legitimate if there are no factors known to modify this effect [77,96,97]. Official guidelines take into account a wider scope of aspects: from the quality of the evidence (when evidence there is), to how to apply the data in day-to-day practice (choices in life-threatening situations, benefit-risk balance, existence of convincing alternatives, cost to society, equity, patients' beliefs or preferences etc.). In this context of guidelines, the GRADE system is relevant [37,68,98].

The foremost aim of REB is to inform the patient and the general practitioners of the existence of this evidence so that an informed choice is available in keeping with EBM [99-100]. Here is the strength of REB: to allow an informed decision even in the absence of a proven effect of the drug. Today, without knowing if evidence is available, whether it is unequivocal or not, choices cannot be informed and so justifications can be implicit, vague and confused. This project will feed into a website to help health-care practitioners discuss drug options with patients and is the focus of the next step of our research team.

## **Conclusion**

What does the E in EBM really stand for? The REB project aims to provide both a systematic method of evaluating drugs and an explicit and practical way of presenting the results so that their interpretation may be more objective. Its final aim is to inform patients and general practitioners on the existence or not of firm evidence: a convincing and credible effect of a drug for a PIO in a specific population. The impact of such an evaluation is potentially major, for patients, care-givers and the community as a whole.

## **Author's contributions**

Rémy Boussageon: conceptualization, writing - original draft, review & editing, validation, supervision; Clara Blanchard: writing - original draft, review & editing, validation, project administration; Elodie Charuel: writing - review & editing; Thibault Menini: writing - review & editing; Florain Naudet: writing - review & editing, methodology; Behrouz Kassai: writing - review & editing; François Gueyffier: writing - review & editing, methodology; Michel Cucherat: writing - review & editing; H  l  ne Vaillant-Roussel: writing - review & editing, methodology, validation, supervision.

## **Funding statement**

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

## **Patient and public involvement**

No patients were directly involved in the design, planning and conception of this study.

## **Disclosure of interest**

The authors have no relationship or interest to disclose which could have direct or potential influence on the work.

## References

- [1] Guyatt G, Cairns J, Churchill D, Cook D; Haynes B, Hirsh J, et al. Evidence-based medicine: a new approach to teaching the practice of medicine. *JAMA* 1992;268:2420-5.
- [2] Greenhalgh T, Howick J, Maskrey N, Evidence Based Medicine Renaissance Group. Evidence based medicine: a movement in crisis? *BMJ* 2014;348:g3725.
- [3] Naudet F, Falissard B, Boussageon R, Healy D. Has evidence-based medicine left quackery behind? *Intern Emerg Med* 2015;10:631-4.
- [4] Ioannidis JP. Evidence-based medicine has been hijacked: a report to David Sackett. *J Clin Epidemiol* 2016;73:82–6.
- [5] McCartney M, Treadwell J, Maskrey N, Lehman R. Making evidence based medicine work for individual patients. *BMJ* 2016;353:i2452.
- [6] Heneghan C, Mahtani KR, Goldacre B, Godlee F, Macdonald H, Jarvies D. Evidence based medicine manifesto for better healthcare: A response to systematic bias, wastage, error and fraud in research underpinning patient care. *Evid Based Med* 2017;22:120-2.
- [7] Lenzer, J. Why we can't trust clinical guidelines. *BMJ* 2013;346:3830.
- [8] Angell M. *The truth about the drug companies: how they deceive us and what to do about it.* 2005. Penguin Random House; New York (USA). ISBN : 0375760946.
- [9] Gotzsche PC. *Deadly medicines and organised crime: how big pharma has corrupted healthcare.* 2013. CRC Press Boca Raton (USA). ISBN: 978-1846198847
- [10] Als-Nielsen B, Chen W, Gluud C, Kjaergard LL. Association of funding and conclusions in randomized drug trials: a reflection of treatment effect or adverse events? *JAMA* 2003;290:921-8.
- [11] Melander H, Ahlqvist-Rastad J, Meijer G, Beermann B. Evidence b(i)ased medicine--selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. *BMJ* 2003;326:1171-3.
- [12] Montori VM, Guyatt G. Corruption of the evidence as threat and opportunity for evidence-based medicine. *Harvard Health Policy Rev* 2007;8:145-5.
- [13] Krumholz HM, Ross JS, Presler AH, Egilman DS. What have we learnt from Vioxx? *BMJ* 2007;334:120-3.
- [14] Vedula SS, Bero L, Scherer RW, Dickersin K. Outcome reporting in industry-sponsored trials of gabapentin for off-label use. *N Engl J Med* 2009;361:1963-71.
- [15] Landefeld CS, Steinman MA. The neurontin legacy--marketing through misinformation and manipulation. *N Engl J Med* 2009;360:103-6.
- [16] Cohen D. Rosiglitazone: what went wrong? *BMJ* 2010;341:c4848.
- [17] Light DW, Lexchin J, Darrow JJ. Institutional corruption of pharmaceuticals and the myth of safe and effective drugs. *J Law Med Ethics* 2013;41:590–600.
- [18] Spielmans GI, Kirsch I. Drug approval and drug effectiveness. *Annu Rev Clin Psychol* 2014;10:741-66.

- [19] Le Noury J, Nardo JM, Healy D, Jureidini J, Raven M, Tufanaru C, Abi-Jaoude E. Restoring Study 329: efficacy and harms of paroxetine and imipramine in treatment of major depression in adolescence. *BMJ* 2015;351:h4320.
- [20] Naci H, Ioannidis JP. How good is "evidence" from clinical studies of drug effects and why might such evidence fail in the prediction of the clinical utility of drugs? *Annu Rev Pharmacol Toxicol* 2015;55:169-89.
- [21] Lundh A, Lexchin J, Mintzes B, Schroll JB, Bero L. Industry sponsorship and research outcome. *Cochrane Database Syst Rev* 2017;2:MR000033.
- [22] Howick J. Exploring the asymmetrical relationship between the power of finance bias and evidence. *Perspect Biol Med* 2019;62:159-87.
- [23] Cosgrove L, Shaughnessy AF, Shaneyfelt T. When is a guideline not a guideline? The devil is in the details. *BMJ Evid Based Med* 2018;23:33-6.
- [24] Torgerson T, Wayant C, Cosgrove L, Akl EA, Checketts J, Dal Re R, et al. Ten years later: a review of the US 2009 institute of medicine report on conflicts of interest and solutions for further reform. *BMJ Evid Based Med* 2022;27:46-54.
- [25] Nejstgaard CH, Bero L, Hróbjartsson A, Jørgensen AW, Jørgensen KJ, Le M, Lundh A. Association between conflicts of interest and favourable recommendations in clinical guidelines, advisory committee reports, opinion pieces, and narrative reviews: systematic review. *BMJ* 2020;371:m4234.
- [26] Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med* 2008;358:252-60.
- [27] Every-Palmer S, Howick J. How evidence-based medicine is failing due to biased trials and selective publication. *J Eval Clin Pract* 2014;20:908-14.
- [28] Kaptchuk TJ. Effect of interpretive bias on research evidence. *BMJ* 2003;326:1453-5.
- [29] Strech D, Tilburt J. Value judgments in the analysis and synthesis of evidence. *J Clin Epidemiol* 2008;61:521-4.
- [30] Djulbegovic B, Kumar A, Magazín A, Schroen AT, Soares H, Hozo I, et al. Optimism bias leads to inconclusive results-an empirical study. *J Clin Epidemiol* 2011;64:583-93.
- [31] McCormack J, Greenhalgh T. Seeing what you want to see in randomised controlled trials: versions and perversions of UKPDS data. United Kingdom prospective diabetes study. *BMJ* 2000;320(7251):1720-3.
- [32] Boutron I, Ravaud P. Misrepresentation and distortion of research in biomedical literature. *Proc Natl Acad Sci USA* 2018;115:2613-9. doi:10.1073/pnas.1710755115
- [33] Boutron I. Spin in scientific publications: a frequent detrimental research practice. *Ann Emerg Med* 2020;75:432-4.
- [34] Khan MS, Lateef N, Siddiqi TJ, Rehman KA, Alnaimat S, Khan SU, et al. Level and prevalence of spin in published cardiovascular randomized clinical trial reports with statistically nonsignificant primary outcomes: a systematic review. *JAMA Netw Open* 2019;2:e192622.
- [35] Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ* 1996;312:71-2.
- [36] Guyatt GH, Naylor D, Richardson WS, Green L, Haynes RB, Wilson MC, et al. What is the best evidence for making clinical decisions? *JAMA* 2000;284:3127-8.
- [37] Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of

- evidence and strength of recommendations. *BMJ* 2008;336:924–6.
- [38] Matthys J. On evidence-based medicine. *Lancet* 2017;390:2245.
- [39] Onakpoya IJ, Thomas ET, Lee JJ, Goldacre B, Heneghan CJ. Benefits and harms of pregabalin in the management of neuropathic pain: a rapid review and meta-analysis of randomised clinical trials. *BMJ Open* 2019;9:e023600.
- [40] Haynes RB, Sackett DL, Guyatt GH, Tugwell P. *Clinical epidemiology: how to do clinical practice research*. 3<sup>rd</sup> ed. SA: Lippincott, Williams and Wilkins; 2012. ISBN: 978-0781745246
- [41] Djulbegovic B, Guyatt G. Evidence-based medicine in times of crisis. *J Clin Epidemiol* 2020;126:164-6.
- [42] Maynié-François C, Tudrej B, Tawil E, Naudet F, Huas C, Pouchain D, et al. Quality of evidence of the efficacy of therapeutic interventions on patient-important outcomes in Cochrane's systematic reviews' abstracts: A survey. *Therapie* 2020;76:403-8.
- [43] Fleming PS, Koletsi D, Ioannidis JP, Pandis N. High quality of the evidence for medical and other health-related interventions was uncommon in Cochrane systematic reviews. *J Clin Epidemiol* 2016;78:34-42.
- [44] Howick J, Koletsi D, Pandis N, Fleming PS, Loef M, Walach H, et al. The quality of evidence for medical interventions does not improve or worsen: a metaepidemiological study of Cochrane reviews. *J Clin Epidemiol* 2020;126:154-9.
- [45] Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? *PLoS Med* 2010;7:e1000326.
- [46] Guyatt GH, Meade MO, Jaeschke RZ, Cook DJ, Haynes RB. Practitioners of evidence based care. Not all clinicians need to appraise evidence from scratch but all need some skills. *BMJ* 2000;320:954-5.
- [47] Boussageon R, Foucher E, Huas C, Tudrej B, Rat C, Pouchain D, et al. Do general practitioners read primary sources of information? *Médecine* 2017;13:378-82.
- [48] Moynihan R. It's time to rebuild the evidence base. *BMJ* 2011;342:d3004.
- [49] Moynihan R, Henry D, Moons KG. Using evidence to combat overdiagnosis and overtreatment: evaluating treatments, tests, and disease definitions in the time of too much. *PLoS Med* 2014;11:e1001655.
- [50] Prasad V, Cifu A, Ioannidis JP. Reversals of established medical practices: evidence to abandon ship. *JAMA* 2012;307:37–8.
- [51] Schwartz LM, Woloshin S. The drug facts box: improving the communication of prescription drug information. *Proc Natl Acad Sci USA* 2013;110:14069–74.
- [52] Stacey D, Légaré F, Lyddiatt A, Giguere AM, Yoganathan M, Saarimaki A, et al. Translating evidence to facilitate shared decision making: development and usability of a consult decision aid prototype. *Patient* 2016;9:571–82.
- [53] Matthys J, De Meyere M, van Driel ML, De Sutter A. Differences among international pharyngitis guidelines: not just academic. *Ann Fam Med* 2007;5:436-43.
- [54] Tudrej BV, Favard D, Vaillant-Roussel H, Pouchain D, Jaafari N, Boussageon R. Validity of the good practice guidelines: the example of type 2 diabetes. *Diabetes Res Clin Pract* 2020;169:108459.
- [55] Oxman AD, Glasziou P, Williams JW Jr. What should clinicians do when faced with conflicting recommendations? *BMJ* 2008;337:a2530.
- [56] Naudet F, Schuit E, Ioannidis JPA. Overlapping network meta-analyses on the same topic: survey of published studies. *Int J Epidemiol*

2017;46:1999-2008.

- [57] Ioannidis JP. The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses. *Milbank Q* 2016;94:485-514.
- [58] Boussageon R, Huas C, Rat C, Pouchain D, Renard V, Supper I, et al. Reconstruire l'Evidence Based pour une décision médicale partagée. Evaluation de l'efficacité et de la sécurité des médicaments fondée sur une revue systématique des essais cliniques randomisés et des méta-analyses. *Exercer* 2017;130:90-4.
- [59] Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ* 2015;350:g7647.
- [60] Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions* version 6.2 (updated February 2021). Cochrane, 2021. <http://www.training.cochrane.org/handbook>. [Accessed 10 February 2021].
- [61] Ranstam J. Hypothesis-generating and confirmatory studies, Bonferroni correction, and pre-specification of trial endpoints. *Acta Orthop* 2019;90(4):297.
- [62] Hennekens CH, DeMets D. The need for large-scale randomized evidence without undue emphasis on small trials, meta-analyses, or subgroup analyses. *JAMA* 2009;302:2361-2.
- [63] European Medicines Agency. ICH Topic E 9 Statistical principles for clinical trials. 1998. [https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-9-statistical-principles-clinical-trials-step-5\\_en.pdf](https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-9-statistical-principles-clinical-trials-step-5_en.pdf). [Accessed 3 October 2022 (37 pp.)].
- [64] Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: A revised tool for assessing risk of bias in randomised trials. *BMJ* 2019;366:l4898.
- [65] Tukey JW. Some thoughts on clinical trials, especially problems of multiplicity. *Science* 1977;198:679-84.
- [66] Bauer P. Multiple testing in clinical trials. *Stat Med* 1991;10:871-90.
- [67] Cucherat M, Laporte S. False positive results or what's the probability that a significant P-value indicates a true effect? *Thérapie* 2017;72:421-6.
- [68] Atkins D, Best D, Briss P, Eccles M, Falck-Ytter Y, Flottorp S, et al. Grading quality of evidence and strength of recommendations. *BMJ* 2004;328:1490.
- [69] Vandembroucke JP, Psaty BM. Benefits and risks of drug treatments. How to combine the best evidence on benefits with the best data about adverse effects. *JAMA* 2008;300: 2417-9.
- [70] Bossard N, Boissel FH, Boissel JP. Level of evidence and therapeutic evaluation: need for more thoughts. *Fundam Clin Pharmacol* 2004;18: 365-72.
- [71] McAlister FA. The Canadian Hypertension Education Program –a unique Canadian initiative. *Can J Cardiol* 2006;22:559-64.
- [72] Garattini S, Jakobsen JC, Wetterslev J, Bertelé V, Banzi R, Rath A, et al. Evidence-based clinical practice: Overview of threats to the validity of evidence and how to minimise them. *Eur J Intern Med* 2016;32:13-21.
- [73] Murad MH, Montori VM. Synthesizing evidence: shifting the focus from individual studies to the body of evidence. *JAMA* 2013;309:2217-8.



- [74] Wetterslev J, Jakobsen JC, Gluud C. Trial sequential analysis in systematic reviews with meta-analysis. *BMC Med Res Methodol* 2017;17:39.
- [75] Murad MH, Mustafa R, Morgan R, Sultan S, Falck-Ytter Y, Dahm P. Rating the quality of evidence is by necessity a matter of judgment. *J Clin Epidemiol* 2016;74:237-8.
- [76] Hopewell S, Boutron I, Altman DG, Ravaud P. Incorporation of assessments of risk of bias of primary studies in systematic reviews of randomised trials: a cross-sectional study. *BMJ Open* 2013;3:e003342.
- [77] Schünemann HJ, Vist GE, Higgins JPT et al. Chapter 15: Interpreting results and drawing conclusions. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions* version 6.2 (updated February 2021). Cochrane, 2021. <http://www.training.cochrane.org/handbook>. [Accessed 4 October 2021].
- [78] Djulbegovic B, Paul A. From efficacy to effectiveness in the face of uncertainty: indication creep and prevention creep. *JAMA* 2011;305:2005-6.
- [79] Janiaud P, Cornu C, Kassai B. Extrapolation will never replace randomized clinical trials. *J Clin Epidemiol* 2016;71:109-10.
- [80] Schuit E, Li AH, Ioannidis JPA. How often can meta-analyses of individual-level data individualize treatment? A meta-epidemiologic study. *Int J Epidemiol* 2019;48:596-608.
- [81] Carlisle B, Federico CA, Kimmelman J. Trials that say "maybe": the disconnect between exploratory and confirmatory testing after drug approval. *BMJ* 2018;360:k959.
- [82] Federico CA, Wang T, Doussau A, Mogil JS, Fergusson D, Kimmelman J. Assessment of pregabalin postapproval trials and the suggestion of efficacy for new indications: a systematic review. *JAMA Intern Med* 2019;179:90-7.
- [83] Agabio R, Trogu E, Pani PP. Antidepressants for the treatment of people with co-occurring depression and alcohol dependence. *Cochrane Database Syst Rev* 2018;4:CD008581.
- [84] Ostuzzi G, Matcham F, Dauchy S, Barbui C, Hotopf M. Antidepressants for the treatment of depression in people with cancer. *Cochrane Database Syst Rev* 2018;4:CD011006.
- [85] Orgeta V, Tabet N, Nilforooshan R, Howard R. Efficacy of antidepressants for depression in Alzheimer's disease: systematic review and meta-analysis. *J Alzheimers Dis* 2017;58:725-33.
- [86] Jakobsen JC, Katakam KK, Schou A, Hellmuth SG, Stallknecht SE, Leth-Møller K, et al. Selective serotonin reuptake inhibitors versus placebo in patients with major depressive disorder. A systematic review with meta-analysis and trial sequential analysis. *BMC Psychiatry* 2017;17:58.
- [87] Raber I, McCarthy CP, Vaduganathan M, Bhatt DL, Wood DA, Cleland JGF, et al. The rise and fall of aspirin in the primary prevention of cardiovascular disease. *Lancet* 2019;393:2155-67.
- [88] Rehman MB, Tudrej BV, Soustre J, Buisson M, Archambault P, Pouchain D, et al. Efficacy and safety of DPP-4 inhibitors in patients with type 2 diabetes: Meta-analysis of placebo-controlled randomized clinical trials. *Diabetes Metab* 2017;43:48-58.
- [89] Bethel MA, Patel RA, Merrill P, Lokhnygina Y, Buse JB, Mentz RJ, et al. Cardiovascular outcomes with glucagon-like peptide-1 receptor agonists in patients with type 2 diabetes: a meta-analysis. *Lancet Diabetes Endocrinol* 2018;6:105-13.
- [90] Liu J, Li L, Deng K, Xu C, Busse JW, Vandvik PO, et al. Incretin based treatments and mortality in patients with type 2 diabetes: systematic review and meta-analysis. *BMJ* 2017;357:j2499.

- [91] Haynes B. Can it work? Does it work? Is it worth it? The testing of healthcare interventions is evolving. *BMJ* 1999;319:652-3.
- [92] Haynes RB, Devereaux PJ, Guyatt GH. Clinical expertise in the era of evidence-based medicine and patient choice. *ACP J Club* 2002;136:A11-4.
- [93] Djulbegovic B. Lifting the fog of uncertainty from the practice of medicine. *BMJ* 2004;329:1419-20.
- [94] Pocock SJ, Stone GW. The primary outcome is positive - is that good enough? *N Engl J Med* 2016;375:971-9.
- [95] Boussageon R, Naudet F, Howick J, Falissard B. Therapeutic efficacy. Between scientific medicine and healthcare medicine. *Presse Med* 2016;45:700-4.
- [96] Dans AL, Dans LF, Guyatt GH, Richardson S, for the Evidence-Based Medicine Working Group. Users' guides to the medical literature: XIV. How to decide on the applicability of clinical trial results to your patient. Evidence-Based Medicine Working Group. *JAMA* 1998;279:545-9.
- [97] Post PN, de Beer H, Guyatt GH. How to generalize efficacy results of randomized trials: recommendations based on a systematic review of possible approaches. *J Eval Clin Pract* 2013;19:638-43.
- [98] Andrews JC, Schünemann HJ, Oxman AD, Pottie K, Meerpohl JJ, Coello PA, et al. GRADE guidelines: 15. Going from evidence to recommendation-determinants of a recommendation's direction and strength. *J Clin Epidemiol* 2013;66:726-35.
- [99] Hoffmann TC, Montori VM, Del Mar C. The connection between evidence-based medicine and shared decision making. *JAMA* 2014;312:1295-6.
- [100] Gigerenzer G, Muir Gray JM, editors. Better doctors, better patients, better decisions. *Envisioning Health Care 2020*. The MIT Press, 2011. <http://www.jstor.org/stable/j.ctt5hhc8b> [Accessed 5 october 2021]

## Legend of figure

**Figure 1.** REB project flow-chart and interpretation table.

### Box 1. Examples of spins: fluctuating interpretations of what is convincing enough.

“The crucial question that making a decision imposes is, is the evidence good enough to act on?” [1]. Evidence is expressed on a continuum scale of credibility, whereas decision making is about choice and is a categorical exercise—we decide or do not [2-4]. So how can we define the level of confidence where a result is solid enough to act upon? Confronted with the same data, from RCTs or from meta-analysis of RCTs, some conclude that the evidence is sound, while others don't. This suggests that the soundness of evidence depends on the context rather than the internal validity of the study. For example, following the publication of the results of UKPDS 34 [5], the medical community along with many promoters of evidence-based medicine (EBM) [6,7], considered that metformin had proven its efficacy for overweight patients with type 2 diabetes, even though the study presented many risks of bias [8]. Also, in *Evidence Based Medicine*, Gerstein [6] concluded that metformin was efficient although the study was not double-blind, although Schulz at the same period highlighted the importance of double blinding to prove drug efficacy [9]. At the time no other treatment had proven its efficacy for type 2 diabetes [5] which probably explains why the results for metformin were deemed convincing. If the results had been published today they would probably have been considered more cautiously and a replication of the results or a confirmatory RCT would have been required. This has not been done to this day [8,10]. This shows how context or implicit preconceived values can bend the interpretation of results [11,12]. Today, metformin remains the gold standard although different meta-analyses of RCTs have concluded that it is not clinically beneficial (no lowering of morbimortality) [8,10,13]. In the same way, when evaluating anti-depressants, a meta-analysis [14] with 82% of low or moderate quality RCTs is unconvincing for some authors [15-17]. For others [14,18] this does not affect the credibility of the results. In another domain, a distinct interpretation of what was convincing [12] led French health-authorities to stop reimbursing anticholinesterase drugs in Alzheimer's disease [19]. The use of steroids for acute bacterial meningitis is also exemplary and well described by Ioannidis [20]. In 2007 a Cochrane meta-analysis of 18 studies concluded that steroids reduced the total risk of mortality (RR = 0.83, 95% CI 0.71 to 0.99) and the rates of severe hearing loss (RR 0.65, 95% CI 0.47 to 0.91) [21]. In 2015, the addition of new studies (25 in total) suggested that these benefits were probably true both for total mortality (RR 0.90, 95% CI 0.80 to 1.01) and for rates of hearing loss (RR 0.67, 95% CI 0.51 to 0.88) [22]. However, if the analysis were restricted to only the 4 studies that were of high quality, then the results were not reversed (mortality: RR 1.00, 95% CI 0.88 to 1.14; severe hearing loss: RR 0.90, 95% CI 0.73 to 1.12) [21]. In practice, can we consider steroids to be efficient or not? Can doctors be entrusted with judging what is convincing data? In the absence of clear cut-off or criteria, EBM has not solved the problem of how to interpret results to allow a transparent therapeutic decision. Some authors [23,24] have justified the prescription of potentially iatrogenic drugs on the basis of poor quality studies or observational studies because it was considered the best available evidence at the time. On the contrary, REB defines evidence, i.e. what is convincing, a priori with a rigorous and objective methodology, so that results are interpreted independently from the context or preconceptions.

[1] Broadbent, Alex. (2013). *Philosophy of Epidemiology*. 10.1057/9781137315601.

- [2] Atkins D, Best D, Briss P, Eccles M, Falck-Ytter Y, Flottorp S, et al. Grading quality of evidence and strength of recommendations. *BMJ* 2004;328:1490.
- [3] Djulbegovic B. Lifting the fog of uncertainty from the practice of medicine. *BMJ* 2004;329:1419-20.
- [4] Djulbegovic B, Guyatt G. Evidence-based medicine in times of crisis. *J Clin Epidemiol* 2020;126:164-6.
- [5] UKPDS Group. Effect of intensive blood-glucose control with metformin on complications in overweight patients with type 2 diabetes (UKPDS 34). UK Prospective Diabetes Study (UKPDS) Group. *Lancet* 1998;352:854-65.
- [6] Gerstein HC. Metformin reduced diabetes-related end points and all-cause mortality in overweight patients with type 2 diabetes. *Evidence Based Medicine* 1999;11.
- [7] Saenz A, Fernandez- Esteban I, Mataix A, Ausejo Segura M, Roqué I, Figuls M, et al. Metformin monotherapy for type 2 diabetes mellitus. *Cochrane Database Syst Rev* 2005; CD002966.
- [8] Boussageon R, Gueyffier F, Cornu C. Metformin as first line treatment for type 2 diabetes: are we sure? *BMJ* 2016;352:h6748.
- [9] Schulz KF. Assessing allocation concealment and blinding in randomised controlled trials: why bother? *Evid Based Nurs.*2001;4:4-6.
- [10] Gnesin F, Thuesen ACB, Kähler LKA, Madsbad S, Hemmingsen B. Metformin monotherapy for adults with type 2 diabetes mellitus. *Cochrane Database Syst Rev* 2020;6:CD012906.
- [11] Kaptchuk TJ. Effect of interpretive bias on research evidence. *BMJ* 2003;326:1453-5.
- [12] Strech D, Tilburt J. Value judgments in the analysis and synthesis of evidence. *J Clin Epidemiol* 2008;61:521-4.
- [13] Boussageon R, Supper I, Bejan-Angoulvant T, Kellou N, Cucherat M, Boissel JP, et al. Reappraisal of metformin efficacy in the treatment of type 2 diabetes: a meta-analysis of randomised controlled trials. *PLoS Med* 2012;9:e1001204.
- [14] Cipriani A, Furukawa TA, Salanti G, Chaimani A, Atkinson LZ, Ogawa Y, et al. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. *Lancet* 2018;391:1357-66.
- [15] McCormack J, Korownyk C. Effectiveness of antidepressants. *BMJ* 2018;360:k1073.
- [16] Jakobsen JC, Gluud C, Kirsch I. Should antidepressants be used for major depressive disorder? *BMJ Evid Based Med* 2020;25:130.
- [17] Munkholm K, Paludan-Müller AS, Boesen K. Considering the methodological limitations in the evidence base of antidepressants for depression: a reanalysis of a network meta-analysis. *BMJ Open* 2019;9:e024886.
- [18] Cipriani A, Salanti G, Furukawa TA, Egger M, Leucht S, Ruhe HG, et al. Antidepressants might work for people with major depression: where do we go from here? *Lancet Psychiatry* 2018;5:461-3.
- [19] Walsh S, King E, Brayne C. France removes state funding for dementia drugs. *BMJ* 2019;367:l6930.
- [20] Ioannidis JP. Meta- research: The art of getting it wrong. *Res Synth Methods* 2010;1: 169-84.
- [21] Van de Beek D, de Gans J, McIntyre P, Prasad K. Corticosteroids for acute bacterial meningitis. *Cochrane Database Syst Rev* 2007;:CD004405.
- [22] Brouwer MC, McIntyre P, Prasad K, van de Beek D. Corticosteroids for acute bacterial meningitis. *Cochrane Database Syst Rev* 2015;:CD004405.

- [23] Nardelli P, Zangrillo A, Sanchini G, Likhvantsev VV, Yavorovskiy AG, Romero Garcia CS, et al. Crying wolf in time of Corona: the strange case of ivermectin and hydroxychloroquine. Is the fear of failure withholding potential life-saving treatment from clinical use? *Signa Vitae* 2021;17:3-4.
- [24] Bryant A, Lawrie TA, Dowsell T, Fordham EJ, Mitchell S, Hill SR, et al. Ivermectin for prevention and treatment of COVID-19 infection: a systematic review, meta-analysis, and trial sequential analysis to inform clinical guidelines. *Am J Ther* 2021;28:e434-e460.

Journal Pre-proof

**Box 2. Can we trust guidelines and marketing authorisations?**

Many drug approvals rest on shaky ground. For example, according to one study, more than half of new drugs entering the German healthcare system have not been shown to add benefit [1]. In France, for example, acetylleucine prescribed for vertigo [2] or phloroglucinol [3] for pelvic pains in gynecology have been approved by health authorities when there is no proof of clinical benefit. Likewise, FDA approvals for new drugs are more and more rarely based on high quality pivotal studies [4], this is particularly true for cancer drugs [5-7]. Even once approved, the use of the medication can drift away from the indication it is approved for (off-label use), in which case there is no proof of efficacy as it is the case for pregabalin in lower back pain or sciatalgia [8-11].

Regarding guidelines, reliability is fickle, as systematic reviews are not always included in their methodology [12]. In a systematic review including 52 international guidelines for the treatment of type 2 diabetes [13], 58% stated they had carried out a systematic review. Only one guideline cited three major meta-analyses published in journals with impact factors superior to 10. Three quarter of the guidelines failed to detail their bibliographic research methods. Also, most were based on low quality evidence (Level of Evidence B ou C) [14,15], including those considering drug efficacy [16]. One such example is the recommended prescription of aspirine in primary prevention of cardiovascular disease in some guidelines [17,18] when confirmatory RCTs have failed to show a clear benefit [19] and 3 megatrials seem to confirm the lack of net benefit [20-22]. Aspirine is also recommended by the USPTF [18] in the prevention of colorectal cancer on the basis of purely exploratory results, with no confirmatory trial to validate its use in this indication. How can guidelines replace primary sources of data in the decision-making process?

- [1] Wieseler B, McGauran N, Kaiser T. New drugs: where did we go wrong and what can we do better? *BMJ* 2019;366:l4340.
- [2] Vanderkam P, Blanchard C, Naudet F, Pouchain D, Vaillant Roussel H, Perault-Pochat MC, et al. Efficacy of acetylleucine in vertigo and dizziness: a systematic review of randomised controlled trials. *Eur J Clin Pharmacol* 2019;75:603-7.
- [3] Blanchard C, Vanderkam P, Pouchain D, Stéphanie M, Hélène VR, Rémy B.. Efficacy of phloroglucinol for the treatment of pain of gynaecologic or obstetrical origin: a systematic review of literature of randomised controlled trials. *Eur J Clin Pharmacol* 2020;76:1-6.
- [4] Zhang AD, Puthumana J, Downing NS, Shah ND, Krumholz HM, Ross JS.. Assessment of clinical trials supporting US Food and Drug Administration approval of novel therapeutic agents, 1995-2017. *JAMA Netw Open* 2020;3:e203284.
- [5] Chen EY, Raghunathan V, Prasad V. An overview of cancer drugs approved by the US Food and Drug Administration based on the surrogate end point of response rate. *JAMA Intern Med* 2019;179:915-21.
- [6] Hilal T, Gonzalez-Velez M, Prasad V. Limitations in clinical trials leading to anticancer drug approvals by the US Food and Drug Administration. *JAMA Intern Med* 2020;180:1108-15.
- [7] Prasad V. Do cancer drugs improve survival or quality of life? *BMJ* 2017;359:j4528.
- [8] Carlisle B, Federico CA, Kimmelman J. Trials that say "maybe": the disconnect between exploratory and confirmatory testing after drug approval. *BMJ* 2018;360:k959.
- [9] Federico CA, Wang T, Doussau A, Mogil JS, Fergusson D, Kimmelman J. Assessment of pregabalin postapproval trials and the suggestion of efficacy for new indications: a systematic review. *JAMA Intern Med* 2019;179:90-7.

- [10] Shanthanna H, Gilron I, Rajarathinam M, AlAmri R, Kamath S, Thabane L, et al. Benefits and safety of gabapentinoids in chronic low back pain: a systematic review and meta-analysis of randomized controlled trials. *PLoS Med* 2017;14:e1002369.
- [11] Enke O, New HA, New CH, Mathieson S, McLachlan AJ, Latimer J, et al. Anticonvulsants in the treatment of low back pain and lumbar radicular pain: a systematic review and meta-analysis. *CMAJ* 2018;190:E786-E793.
- [12] Lunny C, Ramasubbu C, Puil L, Liu T, Gerrish S, Salzwedel DM, et al. Over half of clinical practice guidelines use non-systematic methods to inform recommendations: A methods study. *PLoS One* 2021;16:e0250356.
- [13] Tudrej BV, Favard D, Vaillant-Roussel H, Pouchain D, Jaafari N, Boussageon R. Validity of the good practice guidelines: the example of type 2 diabetes. *Diabetes Res Clin Pract* 2020;169:108459.
- [14] Fanaroff AC, Califf RM, Lopes RD. High-quality evidence to inform clinical practice. *Lancet* 2019;394:633-4.
- [15] Fanaroff AC, Califf RM, Windecker S, Smith SC Jr, Lopes RD. Levels of evidence supporting American College of Cardiology/American Heart Association and European Society of Cardiology guidelines, 2008-2018. *JAMA* 2019;321:1069-80.
- [16] Wagner J, Marquart J, Ruby J, Lammers A, Mailankody S, Kaestner V, Prasad V. Frequency and level of evidence used in recommendations by the National Comprehensive Cancer Network guidelines beyond approvals of the US Food and Drug Administration: retrospective observational study. *BMJ* 2018;360:k668.
- [17] Bibbins-Domingo K, U.S. Preventive Services Task Force. Aspirin use for the primary prevention of cardiovascular disease and colorectal cancer: U.S. Preventive Services Task Force Recommendation Statement. *Ann Intern Med* 2016;164:836-45.
- [18] Fox CS, Golden SH, Anderson C, Bray GA, Burke LE, de Boer IH, et al. Update on prevention of cardiovascular disease in adults with type 2 diabetes mellitus in light of recent evidence: a scientific statement from the American Heart Association and the American Diabetes Association. *Circulation* 2015;132:691-718.
- [19] Raber I, McCarthy CP, Vaduganathan M, Bhatt DL, Wood DA, Cleland JGF, et al. The rise and fall of aspirin in the primary prevention of cardiovascular disease. *Lancet* 2019;393:2155-67.
- [20] ASCEND Study Collaborative Group, Bowman L, Mafham M, Wallendszus K, et al. Effects of aspirin for primary prevention in persons with diabetes mellitus. *N Engl J Med* 2018;379:1529-39.
- [21] McNeil JJ, Wolfe R, Woods RL, Tonkin AM, Donnan GA, Nelson MR, et al. effect of aspirin on cardiovascular events and bleeding in the healthy elderly. *N Engl J Med* 2018;379:1509-18.
- [22] Gaziano JM, Brotons C, Coppolecchia R, Cricelli C, Darius H, Gorelick PB, et al. Use of aspirin to reduce risk of initial vascular events in patients at moderate risk of cardiovascular disease (ARRIVE): a randomised, double-blind, placebo-controlled trial. *Lancet* 2018;392:1036-46.

**Box 3. Confirmatory versus exploratory, why it matters.**

The results of a meta-analysis of exploratory RCTs are only hypothesis generating, like the RCTs it explores. These results are exploratory and cannot be decisive on their own. Confirmatory RCTs are necessary to validate hypotheses and to back decisions with science. They follow the rules of hypothetico-deductive logic (Neyman and Pearson) by defining a priori the hypothesis, the design, the statistical analysis etc. to protect from type I error [1]. Results of confirmatory RCTs rely on the decision logic [2]. In theory, these type of studies are necessary for the European Medicines Agency (EMA) and the United States of America (USA) Food and Drug Administration (FDA) to issue statements for drugs [3]. The results of confirmatory trials is opposed to that of exploratory trials which are not planned a priori and are victims to type 1 error increase. These results are hypothesis-generating [1,4] as are other analyses: sub-group, secondary outcomes or post-hoc (where test multiplicity is not accounted for) [5,6]. Meta-analyses are not safe from false positives results because of the multiplicity of statistical tests [7,8] and because they do not follow the hypothetico-deductive logic [9]. Alone they cannot be decisive and must be confirmed by high quality RCTs. For example a meta-analysis of RCTs tested the efficacy of steroids in the treatment of acute sore throat and pharyngitis[10]. It concluded that steroids were beneficial for “pain at 24hrs”, although all 5 included RCTs were inconclusive for this outcome. Not only that, but one of the studies, of high-quality and testing “pain at 24hrs” as a primary outcome did not find any statistically significant difference (RR = 1,28 ; IC95% = 0,92-1,78;  $p = 0,14$ )[11].The result of the meta-analysis, in favor of the prescription of steroids in this indication, is purely exploratory. Indeed, Hayward et al specify in their methodology that « *all secondary outcomes are considered exploratory* » [11]. How can an exploratory result become decisive in a meta-analysis?

Another example is that of a meta-analysis of RCTs [12] which concluded that vitamine D supplementation lowered the risk of cancer by 16% (with a high level of certainty according to GRADE with a Trial Sequential Analysis) and this, even though no individual RCT measured vitamin D efficacy on risk of cancer as a primary outcome, and one RCT (following 25,000 patients for 5 years) did not yield positive results [13]. According to the REB methodology, the benefit of vitmain D to reduce the risk of cancer remains to be confirmed.

- [1] Ranstam J. Hypothesis-generating and confirmatory studies, Bonferroni correction, and pre-specification of trial endpoints. *Acta Orthop* 2019;90:297.
- [2] Tukey JW. Conclusions vs decisions. *Technometrics* 1960;2:423-33.
- [3] European Medicines Agency. ICH Topic E 9 Statistical Principles for Clinical Trials. 1998. [https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-9-statistical-principles-clinical-trials-step-5\\_en.pdf](https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-9-statistical-principles-clinical-trials-step-5_en.pdf). [Accessed 4 October 2022 (37 pp.)]
- [4] Tukey JW. We need both exploratory and confirmatory. *Am Stat* 1980;34:23-5.
- [5] Tukey JW. Some thoughts on clinical trials, especially problems of multiplicity. *Science* 1977;198:679-84.
- [6] Bender R, Bunce C, Clarke M, Gates S, Lange S, Pace NL, et al. Attention should be given to multiplicity issues in systematic reviews. *J Clin Epidemiol* 2008;61:857-65.
- [7] Ioannidis JP. Meta-research: The art of getting it wrong. *Res Synthesis Methods* 2010;1:169-84.
- [8] Polanin JR, Pigott TD. The use of meta-analytic statistical significance testing. *Res Synth Methods* 2015;6:63-73.

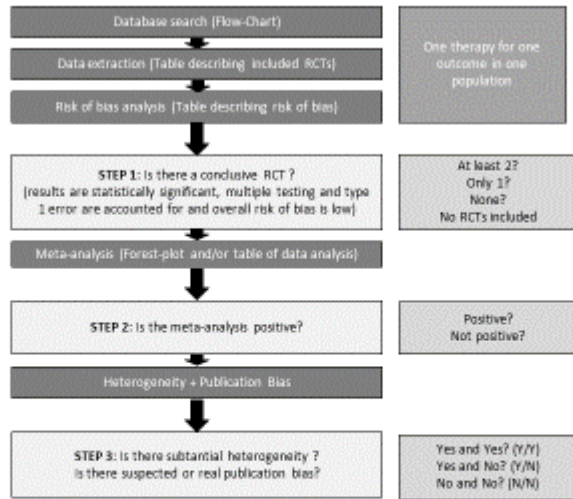


- [9] Vital-Durand D. Problems raised by the conducting and interpretation of meta-analysis. *Therapie* 1994;49:165-8.
- [10] Sadeghirad B, Siemieniuk RAC, Brignardello-Petersen R, Papola D, Lytvyn L, Vandvik PO, et al. Corticosteroids for treatment of sore throat: systematic review and meta-analysis of randomised trials. *BMJ* 2017;358:j3887.
- [11] Hayward GN, Hay AD, Moore MV, Jawad S, Williams N, Voysey M, et al. Effect of oral dexamethasone without immediate antibiotics vs placebo on acute sore throat in adults: a randomized clinical trial. *JAMA* 2017;317:1535-43.
- [12] Zhang Y, Fang F, Tang J, Jia L, Feng Y, Xu P, et al. Association between vitamin D supplementation and mortality: systematic review and meta-analysis. *BMJ* 2019;366:l4673.
- [13] Manson JE, Cook NR, Lee IM, Christen W, Bassuk SS, Mora S, et al. VITAL Research Group. Vitamin D supplements and prevention of cancer and cardiovascular disease. *N Engl J Med* 2019;380:33-44.

**Box 4. Two for tea and why we need two confirmatory trials.**

In theory, the absolute proof of therapeutic effect rests on positive results of at least two confirmatory trials with a low risk of bias [1], with no contradictory results form a meta-analysis of all trials testing the same drug on the same population. This guarantees the stability of the result according to the scientific method with causal inference [2,3]. Indeed, initial results may be refuted by second RCTs and their meta-analysis (for example the results of the CHAOS study of Vitamin E in cardiovascular prevention [4] was refuted by an meta-analysis of the body of evidence [5]). This is why, in theory regulatory bodies like the FDA or the EMA, have at some period demanded at least 2 pivotal studies to warrant marketing authorisation [6]. This is debated [7], as a recent study showed that most marketing authorizations of the FDA have recently been based on only one pivotal study [8]. Indeed, for rare diseases only one study may suffice to authorize market access, in a compromise between an ideal decision basis, the unmet medical needs, and the limited number of patients. Also, carrying out a second study can be difficult for mega-studies such as for cardiovascular prevention when thousands of patients are included [6]. On the one hand, bisoprolol for heart failure with reduced ejection fraction has only been tested correctly in the CIBIS 2 study [9]. On the other hand, other examples show that it is possible to duplicate studies as with aspirine in primary prevention [10], atorvastatine in diabetic patients [11,12] or rosuvastatine in patients with heart failure [13,14]. For symptomatic treatments, RCTs are easier to replicate and meta-analyses can include dozens of RCTs [15,16].

- [1] Bossard N, Boissel FH, Boissel JP. Level of evidence and therapeutic evaluation : need for more thoughts. *Fundam Clin Pharmacol* 2004;18:365-72.
- [2] Fiona F, Wilcox J. Reproducibility of Scientific Results. *The Stanford Encyclopedia of Philosophy* (Winter 2018 Edition), Edward N. Zalta (ed) <https://plato.stanford.edu/cgi-bin/encyclopedia/archinfo.cgi?entry=scientific-reproducibility&archive=win2018>.
- [3] Broadbent A. *Philosophy of epidemiology*. 2013. 10.1057/9781137315601.
- [4] Stephens NG, Parsons A, Schofield PM, Kelly F, Cheeseman K, Mitchinson MJ. Randomised controlled trial of vitamin E in patients with coronary disease: Cambridge Heart Antioxidant Study (CHAOS). *Lancet* 1996;347:781-6.
- [5] Myung SK, Ju W, Cho B, Oh SW, Park SM, Koo BK, et al. Efficacy of vitamin and antioxidant supplements in prevention of cardiovascular disease: systematic review and meta-analysis of randomised controlled trials. *BMJ* 2013;346:f10.
- [6] Temple R. How FDA currently makes decisions on clinical studies. *Clin Trials* 2005;2:276-378.
- [7] Haslam A, Prasad V. Confirmatory trials for drugs approved on a single trial. *Circ Cardiovasc Qual Outcomes* 2019;12:e005494.
- [8] Zhang AD, Puthumana J, Downing NS, Shah ND, Krumholz HM, Ross JS. Assessment of clinical trials supporting US Food and Drug Administration approval of novel therapeutic agents, 1995-2017. *JAMA Netw Open* 2020;3(4):e203284.
- [9] CIBIS 2. The Cardiac Insufficiency Bisoprolol Study II (CIBIS-II): a randomised trial. *Lancet* 1999;353:9-13.
- [10] Raber I, McCarthy CP, Vaduganathan M, Bhatt DL, Wood DA, Cleland JGF, et al. The rise and fall of aspirin in the primary prevention of cardiovascular disease. *Lancet* 2019;393:2155-67.
- [11] Colhoun HM, Betteridge DJ, Durrington PN, Hitman GA, Neil HA, Livingstone SJ, et al. Primary prevention of cardiovascular disease with atorvastatin in type 2 diabetes in the Collaborative Atorvastatin Diabetes Study (CARDS): multicentre randomised placebo-controlled trial. *Lancet* 2004;364:685-96.
- [12] Knopp RH, d'Emden M, Smilde JG, Pocock SJ. Efficacy and safety of atorvastatin in the prevention of cardiovascular end points in subjects with type 2 diabetes: the Atorvastatin Study for Prevention of Coronary Heart Disease Endpoints in non-insulin-dependent diabetes mellitus (ASPEN). *Diabetes Care* 2006;29:1478-85.
- [13] Tavazzi L, Maggioni AP, Marchioli R, Barlera S, Franzosi MG, Latini R, et al. Effect of rosuvastatin in patients with chronic heart failure (the GISSI-HF trial): a randomised, double-blind, placebo-controlled trial. *Lancet* 2008;372:1231-9.
- [14] Kjekshus J, Apetrei E, Barrios V, Böhm M, Cleland JG, Cornel JH, et al. (CORONA). Rosuvastatin in older patients with systolic heart failure. *N Engl J Med* 2007;357:2248-61.
- [15] Cipriani A, Furukawa TA, Salanti G, Chaimani A, Atkinson LZ, Ogawa Y, et al. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. *Lancet* 2018;391:1357-66.
- [16] Busse JW, Wang L, Kamaleldin M, Craigie S, Riva JJ, Montoya L, et al. Opioids for chronic noncancer pain: a systematic review and meta-analysis. *JAMA* 2018;320:2448-60.



STEP 1	At least 2 conclusive RCTs					Only 1 conclusive RCT					0 conclusive RCTs			No included RCTs
STEP 2	Positive			Not positive		Positive			Not positive		Positive	Not positive		NA
STEP 3	N/N	Y/N	Y/Y	N/N	Y/N or Y/Y	N/N	Y/N	Y/Y	N/N	Y/N or Y/Y	N/N	Y/N or Y/Y	N/N or Y/Y or Y/N	NA
<b>Interpretation</b>	<b>Firm evidence</b>	Signal	Lack	Signal	Lack	<b>Evidence</b>	Signal	Lack	Signal	Lack	Signal	Lack	Lack	Lack