European Journal of Internal Medicine 32 (2016) 13–21

Contents lists available at ScienceDirect

European Journal of Internal Medicine

journal homepage: www.elsevier.com/locate/ejim

Review Article

# Evidence-based clinical practice: Overview of threats to the validity of evidence and how to minimise them

Silvio Garattini [a], Janus C. Jakobsen [b,c], Jørn Wetterslev [b], Vittorio Bertelé [a], Rita Banzi [a], Ana Rath [d], Edmund A.M. Neugebauer [e], Martine Laville [f], Yvonne Masson [f], Virginie Hivert [f], Michaela Eikermann [g], Burc Aydin [h], Sandra Ngwabyt [d], Cecilia Martinho [i], Chiara Gerardi [a], Cezary A. Szmigielski [j], Jacques Demotes-Mainard [k], Christian Gluud [b,*]

[a] IRCCS Istituto di Ricerche Farmacologiche Mario Negri, Milan, Italy
[b] The Copenhagen Trial Unit, Centre for Clinical Intervention Research, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark
[c] Department of Cardiology, Holbæk Hospital, Holbæk, Denmark
[d] Orphanet, Institut National de la Santé et de la Recherche Médicale US14, Paris, France
[e] Faculty of Health, School of Medicine, Witten/Herdecke University, Campus Cologne, Germany
[f] Centre de Recherche en Nutrition Humaine, Rhône-Alpes, Univ de Lyon, Lyon, France
[g] Institute for Research in Operative Medicine, Faculty of Health, School of Medicine, Witten/Herdecke University, Cologne, Germany
[h] Department of Medical Pharmacology, School of Medicine, Dokuz Eylul University, Izmir, Turkey
[i] Palliative Care Service, Portuguese Institute of Oncology, Porto, Portugal
[j] Department of Internal Medicine, Hypertension and Vascular Diseases, Medical University of Warsaw, Warsaw, Poland
[k] ECRIN (European Clinical Research Infrastructure Network), Paris, France

ARTICLE INFO

*Article history:*
Received 15 March 2016
Accepted 22 March 2016
Available online 6 May 2016

*Keywords:*
Evidence-based medicine
Evidence-based clinical practice
Systematic review
Randomised clinical trial
Meta-analysis
Trial Sequential Analysis

ABSTRACT

Using the best quality of clinical research evidence is essential for choosing the right treatment for patients. How to identify the best research evidence is, however, difficult. In this narrative review we summarise these threats and describe how to minimise them. Pertinent literature was considered through literature searches combined with personal files. Treatments should generally not be chosen based only on evidence from observational studies or single randomised clinical trials. Systematic reviews with meta-analysis of all identifiable randomised clinical trials with Grading of Recommendations Assessment, Development and Evaluation (GRADE) assessment represent the highest level of evidence. Even though systematic reviews are trust worthier than other types of evidence, all levels of the evidence hierarchy are under threats from systematic errors (bias); design errors (abuse of surrogate outcomes, composite outcomes, etc.); and random errors (play of chance). Clinical research infrastructures may help in providing larger and better conducted trials. Trial Sequential Analysis may help in deciding when there is sufficient evidence in meta-analyses. If threats to the validity of clinical research are carefully considered and minimised, research results will be more valid and this will benefit patients and heath care systems.



## 1. Introduction

James Lind conducted his controlled clinical trial on interventions for scurvy in 1747 and since then evidence-based medicine has undergone a fascinating development [1–4]. Before 1900, only a few controlled clinical trials and randomised clinical trials (RCTs) were launched. During the last century, the conduct of RCTs increased importantly and meta-analyses were introduced [1–4].

Regarding medicinal products, an international consensus has been established allowing a phased assessment of intervention effects (Table 1). Certain fields like cardiology and oncology are fortunate to produce large numbers of RCTs [5]. Other fields like neurology, nephrology, endocrinology, hepatology, and surgery are less fortunate [5]. Medical devices, nutrition, and rare diseases are considered fields especially

* Corresponding author at: The Copenhagen Trial Unit, Centre for Clinical Intervention Research, Department 7812, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark. Tel.: +45 40 40 11 82.
*E-mail addresses:* silvio.garattini@marionegri.it (S. Garattini), jcj@ctu.dk (J.C. Jakobsen), wetterslev@ctu.dk (J. Wetterslev), vittorio.bertele@marionegri.it (V. Bertelé), rita.banzi@marionegri.it (R. Banzi), ana.rath@inserm.fr (A. Rath), Edmund.Neugebauer@uni-wh.de (E.A.M. Neugebauer), martine.laville@chu-lyon.fr (M. Laville), yvonne.masson@clermont.inra.fr (Y. Masson), virginie.hivert@inserm.fr (V. Hivert), michaela.eikermann@uni-wh.de (M. Eikermann), burcaydin@gmail.com (B. Aydin), sandra-nadia.ngwabyt-bikeye@inserm.fr (S. Ngwabyt), cvm@aibili.pt (C. Martinho), chiara.gerardi@marionegri.it (C. Gerardi), cezary.szmigielski@wum.edu.pl (C.A. Szmigielski), jacques.demotes@ecrin.org (J. Demotes-Mainard), cgluud@ctu.dk (C. Gluud).

http://dx.doi.org/10.1016/j.ejim.2016.03.020
0953-6205/

**Table 1**
The phases of clinical research regarding preventive or therapeutic medical interventions.

| Phases | Participants and study designs for preventive or therapeutic interventions |
|---|---|
| Phase I | Healthy participants or patients<br>– observational studies<br>– randomised clinical trials<br>designed to assess the safety (pharmacovigilance), tolerability, pharmacokinetics, and pharmacodynamics of an intervention. |
| Phase II | Patients with disease in question<br>– randomised clinical trials.<br>Phase II trials are performed on larger groups (up to about 300 patients) and are designed to continue safety assessments and to assess how well the intervention works. |
| Phase III | Patients with disease in question<br>– randomised clinical trials<br>often multicentre trials on large patient groups (300 to 10,000 or more depending upon the disease and outcome studied) aimed at being the definitive assessment of how effective the intervention is, in comparison with current 'gold standard' treatment. |
| Phase IV | Patients with disease in question<br>– randomised clinical trials<br>– observational studies.<br>These studies and trials study the impact of applying the new intervention in clinical practice. This includes large randomised clinical trials, cluster randomised trials, and observational studies (clinical databases). |

For medical devices slightly different phases are described [104].

in need of better clinical research [5,6]. The European Clinical Research Infrastructures Network (ECRIN)-Integrating Activity (IA) (http://www.ecrin.org/en/cooperative-projects/ecrin-integrating-activity-clinical-research-in-europe) has therefore identified barriers for good clinical research within these fields and assessed how these barriers could be broken down in order to improve their evidence-based clinical practice [7–10].

As an integral part of these activities, we provide an overview of the hierarchy of evidence regarding interventions and consider the threats to the validity of results of RCTs and systematic reviews with meta-analyses. The threats encompass risks of systematic errors ('bias'); design errors (erroneous selection of patients, doses of medication, comparators, analyses, outcomes, etc.); and risks of random errors (misleading results due to 'play of chance') [11–16]. We suggest possible solutions to the threats including establishment of national or transnational research infrastructures like ECRIN to improve clinical research and hereby reduce research waste [17–25].

## 2. Search strategy and selection criteria

Data for this review were identified by searches of PubMed and The Cochrane Library, references from relevant articles using the search terms "evidence based clinical practice", "evidence based medicine", "evidence hierarchy", "bias risks", "design errors", and "random errors", plus personal literature files. Articles were selected with a view that they should represent important didactic efforts to increase the medical profession's understanding of the central importance that evidence quality plays in underpinning clinical practice.

## 3. The hierarchy of evidence

Different experimental designs have different inferential powers, hence the hierarchy of evidence (Fig. 1) [13]. Provided the methodological quality of your study is good, the higher your study is in the hierarchy, the more likely you observe something close to the 'truth'. With better inferential powers, the higher the likelihood for improving patient outcomes when one translates the research findings into clinical practice (TRIP) [13]. All levels of the hierarchy may be threatened by systematic errors; design errors; and random errors [11,13,26].

### 3.1. Systematic reviews and meta-analyses

The Cochrane Collaboration coined the word 'systematic review' back in 1993, and developed The Cochrane Handbook for Systematic Reviews of Interventions (http://www.cochrane.org/training/cochrane-handbook) [11]. Systematic reviews are based upon peer-reviewed protocols and follow standardised methodologies [5,11,27]. Meta-analyses conducted without a protocol run the risk of systematic, design, and random errors, which may cloud our judgement on benefits and harms of interventions, and makes it difficult to design future trials validly [26,28–30].

### 3.2. Systematic reviews with meta-analysis of several small RCTs compared to a single, large RCT

A heated debate about which is superior — the results of a single large RCT or the results of a systematic review of all trials on a given intervention — has been on-going since meta-analyses became widely known in the 1980s. Some claim that evidence produced in a large RCT is much more valuable than results of systematic reviews or meta-analyses [31–33]. The trial advocates consider that systematic reviews should only be viewed as hypothesis-generating research [31–33].

Systematic reviews with meta-analyses cannot always be conducted with the same scientific cogency as a RCT with pre-defined high-quality methodology, addressing an a priori hypothesised intervention effect [11,30]. Systematic review authors will often know some of the RCTs before they have prepared their protocol for the systematic review, and hence, the review methodology will be at least partly data driven [11,30]. Understanding the inherent methodological limitations of systematic reviews with consideration and implementation of an improved review methodology already at the protocol stage can minimise this limitation [30]. Hence, a cornerstone of a high quality systematic review is the application of transparent, rigorous, and reproducible methodology [34].

IntHout and colleagues used simulations to evaluate error proportions in conventionally powered RCTs (80% or 90% power) compared to random-effects model meta-analyses of smaller trials (30% or 50% power) [35]. When a treatment was assumed to have no effect and heterogeneity was present, the errors for a single trial were increased more than 10-fold above the nominal rate, even for low heterogeneity [35]. Conversely, the error rates in meta-analyses were correct [35]. Evidence from a well-conducted systematic review of several RCTs with low risk of bias therefore represents a higher level of evidence compared to the results from a single RCT [11–14,29,30]. It also appears intuitively evident that inclusion of all available data from all RCTs with low risks of bias ever conducted, should be treated as a higher level of evidence compared to the data from one single RCT [13,30].

As a relatively new approach, network meta-analyses allow comparing interventions that have never been tested head to head in RCTs [36]. Careful consideration is needed for network meta-analyses to avoid false positive results [37]. Statistical and conceptual heterogeneity of the trials combined in a network meta-analysis should be assessed to avoid incoherence and thus chance findings [36]. Reporting bias can affect the findings of a network meta-analysis and lead to incorrect conclusions about the treatments compared [38]. Due to high number of pairwise comparisons in a network analysis, the risk of type I error should be controlled (see below). To address these methodological limitations in a systematic way, a clear protocol and a concise hypothesis are needed in advance to justify the meta-analytic approach [37,39].

In order to improve the systematic review methodology, recent PRISMA guidelines have been developed for individual participant data (IPD) systematic reviews with meta-analysis [40] and for network meta-analyses [39].