LIGHTS

Library of Guidance for Health Scientists

www.lights.science

Stefan Schandelmaier,
Universität Basel

Wer eine Studie macht, trifft viele methodische Entscheidungen.

Study design
Literature search
Eligibility criteria
Intervention & comparison
Sample size
Sampling
Randomisation
Selection of outcomes
Outcome measurement
Main confounders
Main statistical analysis
Model selection

Subgroup analysis
Sensitivity analysis
Missing data
Multiple testing
Thresholds for decision making
Certainty of evidence
Sensitivity analysis
Saturation
Patient involvement
Data sharing
Implementation
...

Pirosca *et al. Trials* (2022) 23:458
https://doi.org/10.1186/s13063-022-06415-5

Trials

COMMENTARY Open Access

Tolerating bad health research: the continuing scandal

Stefania Pirosca[1], Frances Shiely[2,3], Mike Clarke[4] and Shaun Treweek[1*]

Abstract

Background: At the 2015 REWARD/EQUATOR conference on research waste, the late Doug Altman revealed that his only regret about his 1994 *BMJ* paper 'The scandal of poor medical research' was that he used the word 'poor' rather than 'bad'. But how much research is bad? And what would improve things?

Main text: We focus on randomised trials and look at scale, participants and cost. We randomly selected up to two quantitative intervention reviews published by all clinical Cochrane Review Groups between May 2020 and April 2021. Data including the risk of bias, number of participants, intervention type and country were extracted for all trials included in selected reviews. High risk of bias trials was classed as bad. The cost of high risk of bias trials was estimated using published estimates of trial cost per participant.

We identified 96 reviews authored by 546 reviewers from 49 clinical Cochrane Review Groups that included 1659 trials done in 84 countries. Of the 1640 trials providing risk of bias information, 1013 (62%) were high risk of bias (bad), 494 (30%) unclear and 133 (8%) low risk of bias. Bad trials were spread across all clinical areas and all countries. Well over 220,000 participants (or 56% of all participants) were in bad trials. The low estimate of the cost of bad trials was £726 million; our high estimate was over £8 billion.

We have five recommendations: trials should be neither funded (1) nor given ethical approval (2) unless they have a statistician and methodologist; trialists should use a risk of bias tool at design (3); more statisticians and methodologists should be trained and supported (4); there should be more funding into applied methodology research and infrastructure (5).

Conclusions: Most randomised trials are bad and most trial participants will be in one. The research community has tolerated this for decades. This has to stop: we need to put rigour and methodology where it belongs — at the centre of our science.

Keywords: Randomised trials, Research waste, Risk of bias, Statisticians, Methodologists

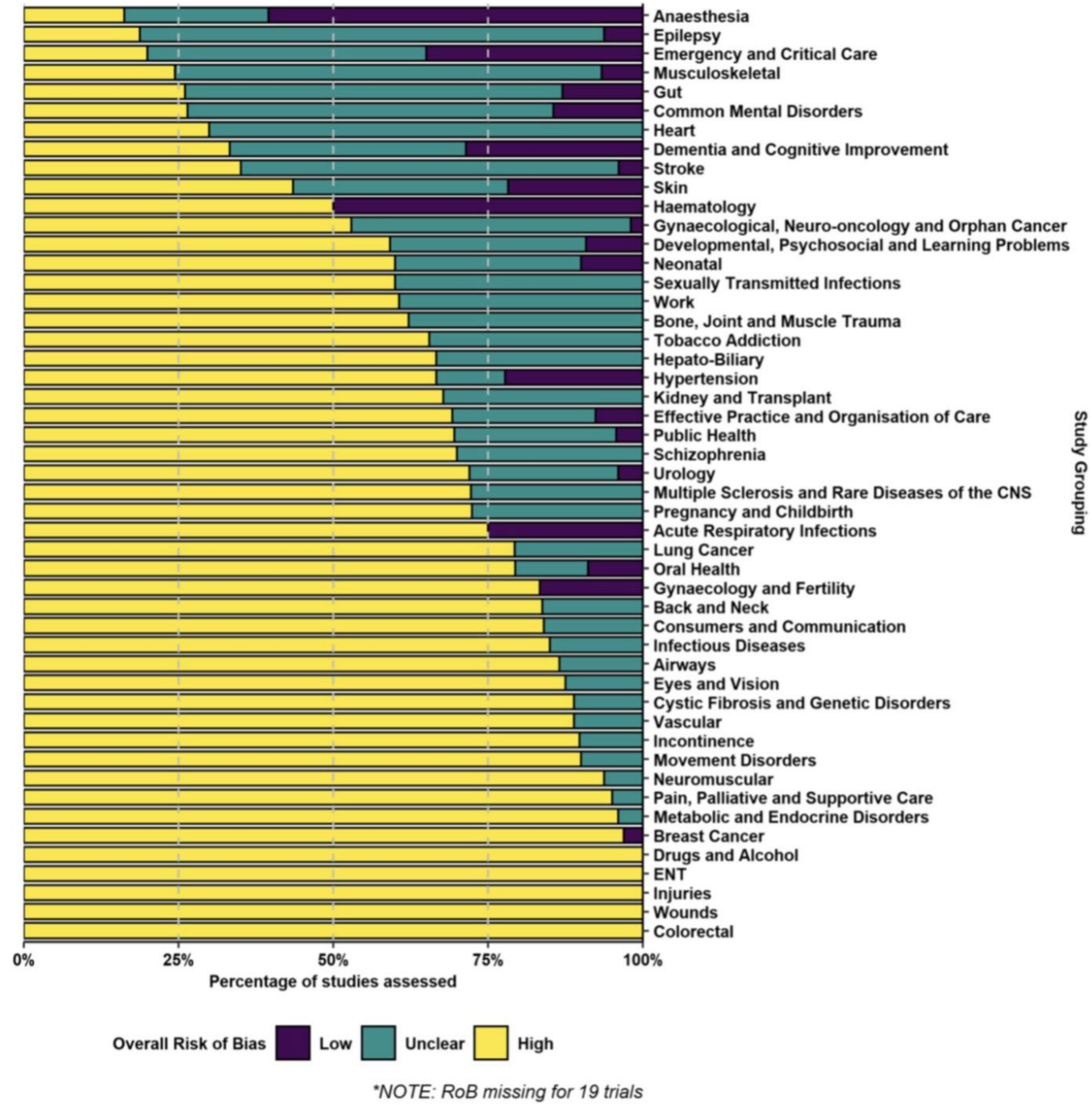
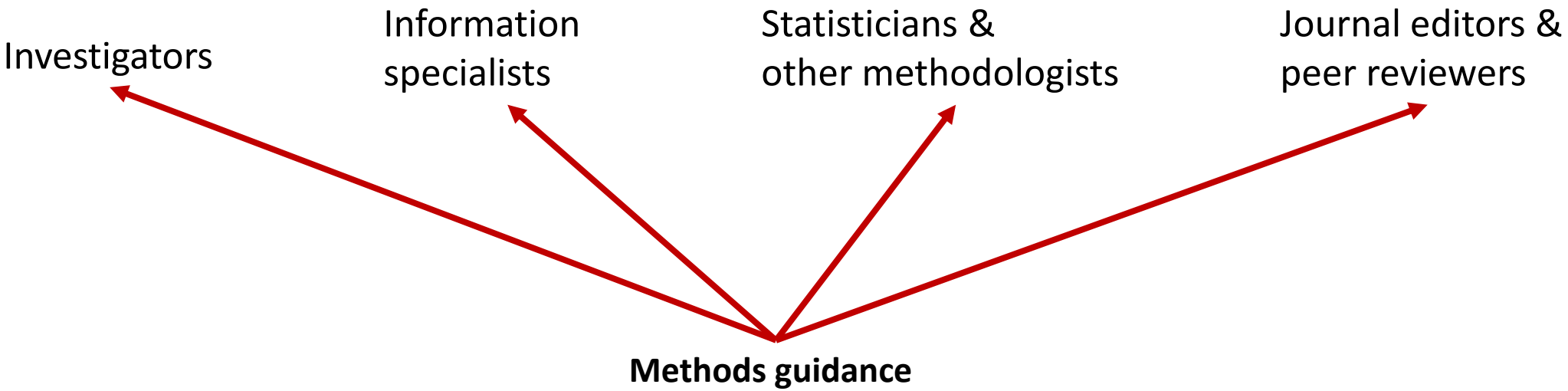


Fig. 1 Risk of bias for included trials in randomly selected systematic reviews published between May 2020 and April 2021 by 49 Cochrane Review Groups


Investigators
Information specialists
Statisticians & other methodologists
Journal editors & peer reviewers
Methods guidance

Received: 9 December 2018 | Revised: 25 February 2020 | Accepted: 28 February 2020

DOI: 10.1002/sim.8532

TUTORIAL IN BIOSTATISTICS

Statistics in Medicine WILEY

STRATOS guidance document on measurement error and misclassification of variables in observational epidemiology: Part 1—Basic theory and simple methods of adjustment

Ruth H. Keogh[1] | **Pamela A. Shaw**[2] | **Paul Gustafson**[3] | **Raymond J. Carroll**[4,5] | **Veronika Deffner**[6] | **Kevin W. Dodd**[7] | **Helmut Küchenhoff**[8] | **Janet A. Tooze**[9] | **Michael P. Wallace**[10] | **Victor Kipnis**[11] | **Laurence S. Freedman**[12,13]

[1]Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, UK

[2]Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, USA

[3]Department of Statistics, University of British Columbia, Vancouver, British Columbia, Canada

[4]Department of Statistics, Texas A&M University, College Station, Texas, USA

[5]School of Mathematical and Physical Sciences, University of Technology Sydney, Broadway, New South Wales, Australia

[6]Statistical Consulting Unit StaBLab, Department of Statistics, Ludwig-Maximilians-Universität, Munich, Germany

[7]Biometry Research Group, Division of Cancer Prevention, National Cancer Institute, Bethesda, Maryland, USA

[8]Department of Statistics, Statistical Consulting Unit StaBLab, Ludwig-Maximilians-Universität, Munich, Germany

[9]Department of Biostatistics and Data Science, Wake Forest School of Medicine, Winston-Salem, North Carolina, USA

[10]Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada

Measurement error and misclassification of variables frequently occur in epidemiology and involve variables important to public health. Their presence can impact strongly on results of statistical analyses involving such variables. However, investigators commonly fail to pay attention to biases resulting from such mismeasurement. We provide, in two parts, an overview of the types of error that occur, their impacts on analytic results, and statistical methods to mitigate the biases that they cause. In this first part, we review different types of measurement error and misclassification, emphasizing the classical, linear, and Berkson models, and on the concepts of nondifferential and differential error. We describe the impacts of these types of error in covariates and in outcome variables on various analyses, including estimation and testing in regression models and estimating distributions. We outline types of ancillary studies required to provide information about such errors and discuss the implications of covariate measurement error for study design. Methods for ascertaining sample size requirements are outlined, both for ancillary studies designed to provide information about measurement error and for main studies where the exposure of interest is measured with error. We describe two of the simpler methods, regression calibration and simulation extrapolation (SIMEX), that adjust for bias in regression coefficients caused by measurement error in continuous covariates, and illustrate their use through examples drawn from the Observing Protein and Energy (OPEN) dietary validation study. Finally, we review software available for implementing these methods. The second part of the article deals with more advanced topics.

KEYWORDS

Berkson error, classical error, differential error, measurement error, misclassification, nondifferential error, regression calibration, sample size, SIMEX, simulation extrapolation

Tutorial, guidance
Abstract unstructured
Development unclear

ELSEVIER

Journal of Clinical Epidemiology 121 (2020) 62–70

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

GRADE Guidelines 28: Use of GRADE for the assessment of evidence about prognostic factors: rating certainty in identification of groups of patients with different absolute risks

Farid Foroutan[a,b,*], Gordon Guyatt[a], Victoria Zuk[a], Per Olav Vandvik[c], Ana Carolina Alba[b], Reem Mustafa[d], Robin Vernooij[e], Ingrid Arevalo-Rodriguez[f,g], Zachary Munn[h], Pavel Roshanov[i], Richard Riley[j], Stefan Schandelmaier[a], Ton Kuijpers[k], Reed Siemieniuk[a], Carlos Canelo-Aybar[e], Holger Schunemann[a], Alfonso Iorio[a]

[a]*Department of Health Research Methods, Evidence, and Impact, McMaster University, Ontario, Canada*
[b]*Ted Rogers Center for Heart Research, Toronto General Hospital, Ontario, Canada*
[c]*Division Gjøvik, Department of Medicine, Innlandet Hospital Trust, Gjøvik, Norway*
[d]*Department of Internal Medicine, The University of Kansas Medical Center, Kansas City, USA*
[e]*Iberoamerican Cochrane Centre, Biomedical Research Institute Sant Pau (IIB Sant Pau), Barcelona, Spain*
[f]*Clinical Biostatistics Unit, Hospital Universitario Ramón y Cajal, IRYCIS, CIBER of Epidemiology and Public Health, Madrid, Spain*
[g]*Centro de investigación en Salud Pública y Epidemiología Clínica. Facultad de Ciencias de la Salud Eugenio Espejo, Universidad UTE, Quito, Ecuador*
[h]*Faculty of Health and Medical Sciences, Joanna Briggs Institute, The University of Adelaide, Adelaide, Australia*
[i]*Department of Medicine, McMaster University, Ontario, Canada*
[j]*School of Health and Population Sciences, University of Birmingham, Birmingham, UK*
[k]*Department of Guideline Development, Dutch College of General Practitioners, Utrecth, The Netherlands*

Accepted 13 December 2019; Published online 23 January 2020

Abstract

Objective: The objective of this study was to provide guidance on the use of the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach to determine certainty in estimates of association between prognostic factors and future outcomes.

Study Design and Setting: We developed our guidance through an iterative process that involved review of published systematic reviews and meta-analyses of prognostic factors, consultation with members, feedback, presentation, and discussion at the GRADE Working Group meetings.

Results: For questions of prognosis, a body of observational evidence (potentially including patients enrolled in randomized controlled trials) begins as high certainty in the evidence. The five domains of GRADE for rating down certainty in the evidence, that is, risk of bias, imprecision, inconsistency, indirectness, and publication bias, as well as the domains for rating up, also apply to estimates of associations between prognostic factors and outcomes. One should determine if their ratings do not consider (noncontextualized) or consider (contextualized) the clinical context as this will may result in variable judgments on certainty of the evidence.

Conclusions: The same principles GRADE proposed for bodies of evidence addressing treatment and overall prognosis work well in assessing individual prognostic factors, both in noncontextualized and contextualized settings.

Keywords: GRADE; Certainty in evidence; Prognosis; Prognostic factor; Guideline; Systematic review; Subgroup

Funding: No external funding.
Ethical approval: Not applicable.
Competing interest statement: All authors declare they did not receive support from any organization for the submitted work; no financial relationships with any organizations that might have an interest in the submitted work in the previous 3 years exist nor do other relationships or activities that could appear to have influenced the submitted work. All authors are members of the GRADE working group.

Authors' contributions: All authors contributed to the generation of the research hypothesis, participated to the discussion of its content, and approved the final version of the article. F.F., V.Z., and A.I. selected the systematic reviews used as examples and prepared summary of finding tables used in the process. F.F. drafted the article, and A.I. is the guarantor.
Data sharing: No additional data available.
* Corresponding author. Tel.: +1 416-340-3482; fax: +1 416-340-4134.
E-mail address: foroutaf@mcmaster.ca (F. Foroutan).

https://doi.org/10.1016/j.jclinepi.2019.12.023

Guideline
Structured abstract
Explicit development process

Treweek *et al. Trials* (2020) 21:33
https://doi.org/10.1186/s13063-019-3980-5

Trials

METHODOLOGY Open Access

Trial Forge Guidance 2: how to decide if a further Study Within A Trial (SWAT) is needed

Shaun Treweek[1*], Simon Bevan[2], Peter Bower[3], Matthias Briel[4], Marion Campbell[1], Jacquie Christie[5], Clive Collett[6], Seonaidh Cotton[1], Declan Devane[7], Adel El Feky[1], Sandra Galvin[7], Heidi Gardner[1], Katie Gillies[1], Kerenza Hood[8], Jan Jansen[9], Roberta Littleford[10], Adwoa Parker[11], Craig Ramsay[1], Lynne Restrup[12], Frank Sullivan[13], David Torgerson[11], Liz Tremain[2], Erik von Elm[14], Matthew Westmore[2], Hywel Williams[15], Paula R. Williamson[16] and Mike Clarke[17]

Abstract

The evidence base available to trialists to support trial process decisions—e.g. how best to recruit and retain participants, how to collect data or how to share the results with participants—is thin. One way to fill gaps in evidence is to run Studies Within A Trial, or SWATs. These are self-contained research studies embedded within a host trial that aim to evaluate or explore alternative ways of delivering or organising a particular trial process. SWATs are increasingly being supported by funders and considered by trialists, especially in the UK and Ireland. At some point, increasing SWAT evidence will lead funders and trialists to ask: given the current body of evidence for a SWAT, do we need a further evaluation in another host trial? A framework for answering such a question is needed to avoid SWATs themselves contributing to research waste.

This paper presents criteria on when enough evidence is available for SWATs that use randomised allocation to compare different interventions.

Introduction

The evidence available to inform many routine process decisions in randomised trials is thin or weak. This includes the evidence on how best to recruit participants [1], retain them [2], collect their data [3] or include them in decisions about the trial [4]. While evidence gaps in, say, the clinical management of diabetes might be expected to lead to a sustained and substantial research effort to fill them, similar effort has not materialised for trial methods research. Recruitment remains a major concern [5, 6] despite more than 25,000 new trials opening every year and needing to recruit participants [7]. Once recruited, there is also little evidence available to inform decisions about how to encourage trial participants to remain in the trial and, for example, to attend face-to-face measurement visits, which are a vital part of most trials [2]. Further, there is almost no evidence base to inform trial management decisions, including how to select sites, whether visiting them in person is worth it, or how to train staff [8].

The lack of trial process evidence contributes to research waste—for example, through poor recruitment, retention and data quality—and has been a feature of medical research for decades [9], with some suggesting that up to 85% of medical research spending is wasted [10]. However, much of the waste is avoidable [11] and research funders recognise the need to avoid it [12].

Trial Forge (http://www.trialforge.org) is an initiative that aims to improve the efficiency of trials, particularly by filling gaps in trial process evidence [13]. One way of improving the evidence base for trial process decisions is to do a Study Within A Trial (SWAT) [14], which is a '...self-contained research study that has been embedded within a host trial with the aim of evaluating or exploring alternative ways of delivering or organising a particular trial process' [15]. For example, a SWAT could evaluate a new way of presenting information to potential participants as

* Correspondence: streweek@mac.com
[1]Health Services Research Unit, University of Aberdeen, Aberdeen, UK
Full list of author information is available at the end of the article

Guidance, decision aid
Abstract unstructured
Development unclear

Monsalves *et al. BMC Medical Research Methodology* (2020) 20:3
https://doi.org/10.1186/s12874-019-0876-8

BMC Medical Research Methodology

TECHNICAL ADVANCE Open Access

LEVEL (Logical Explanations & Visualizations of Estimates in Linear mixed models): recommendations for reporting multilevel data and analyses

Maria Jose Monsalves[1], Ananta Shrikant Bangdiwala[2], Alex Thabane[3] and Shrikant Ishver Bangdiwala[3,4,5,6*]

Keywords: Multilevel models, Reporting guidelines, Variance partition coefficients, Multilevel diagram

Background

Researchers have been utilizing linear mixed models (LMMs) for different hierarchical study designs and under different names, which emphasizes the need for a standard in reporting such models [1, 2]. Mixed effects models, multilevel data, contextual analysis, hierarchical studies, longitudinal studies, panel data and repeated-measures designs are some of the different names used when referring to study designs and/or analytical tools for correlated data. In addition, there is usually no distinction made between having a data structure that is multilevel, and having a research question that requires a multilevel analysis. There are multiple excellent tutorials on multilevel analyses [3–5]. However, there is inconsistency in how the results of LMMs are reported in the literature [6]. Casals et al. conducted a systematic review of how various LMMs were reported in the medical literature, and found that important aspects were not reported in most cases [6].

As an example, a cohort study of children that selects a sample of schools, then selects students within schools, and conducts multiple measurements over time in the same students, would be a 3-level dataset: with school as the highest level (Level 3), student as a lower level (Level 2), and time-point as the lowest level (Level 1). Repeated measurements of a variable over time within a student are likely to be similar, i.e. positively correlated. Also, values of a variable measured on students of a particular school may be more similar to each other than to the values of the same variable measured on students from different schools, i.e. they are also likely to be positively correlated. These within-level correlations reduce the overall information in the data. Considering the correlations typically leads to larger estimates of variances and consequently lower power if sample sizes are not increased at the design stage. At the analysis stage, incorporating random effects into a regression model is one way to acknowledge the variation among upper-level units. Random intercepts and random slopes help to attribute the variation in values of the outcome variable to the relevant levels and independent variables.

A standardized checklist for the reporting of multilevel data and the presentation of linear mixed models will promote adequate reporting of correlated data analyses. In this manuscript, we propose LEVEL (Logical Explanations & Visualizations of Estimates in Linear mixed models), a systematic approach for the presentation of studies with correlated data from multilevel study designs, with an accompanying checklist for standardizing the reporting of results from linear mixed models. These models are quite complex, and the intention of this manuscript is not to be a statistical tutorial, but to mention aspects of the study design and analysis methods that we propose should be addressed in a publication. We present the basics of a linear mixed model simply to introduce the terminology and to help understand the proposed reporting recommendations.

Methods

The linear mixed model

Written as an equation, the 'null' (no covariate) linear mixed model for a 2-level hierarchical study is:

* Correspondence: kant@unc.edu
[3]McMaster University, Hamilton, ON, Canada
[4]Population Health Research Institute, Hamilton, ON, Canada
Full list of author information is available at the end of the article

Recommendations for reporting
No abstract
Development unclear

ELSEVIER

Journal of Clinical Epidemiology ■ (2022) ■

Journal of Clinical Epidemiology

REVIEW ARTICLE

A systematic survey of methods guidance suggests areas for improvement regarding access, development, and transparency

Julian Hirt[a,b], Hannah Ewald[c], Daeria O. Lawson[d], Lars G. Hemkens[a,e,f], Matthias Briel[a,d], Stefan Schandelmaier[a,d,*]

[a]*Department of Clinical Research, University Hospital Basel, University of Basel, Basel, Switzerland*
[b]*International Graduate Academy, Institute for Health and Nursing Science, Medical Faculty, Martin Luther University Halle-Wittenberg, Halle (Saale), Germany*
[c]*University Medical Library, University of Basel, Basel, Switzerland*
[d]*Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Canada*
[e]*Meta-Research Innovation Center Berlin (METRIC-B), Berlin Institute of Health, Berlin, Germany*
[f]*Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Stanford, CA, USA*

Accepted 15 May 2022; Published online xxxx

Abstract

Background: To assess the current practice of developing and presenting methods guidance and explore opportunities for improvement.

Systematic survey of 105 methods guidance articles published in 12 influential journals in 2020

- 12% reporting guidance / 88% other guidance
- 38 different expressions for guidance
- 38% had MeSH terms, 17% author keywords expressing guidance
- 42% Structured abstract
- 42% Development process reported

Databases for methods guidance

Planning ?

Conduct ?

Analysis ?

Reporting

Interpretation ?

Quality assessment ?

… ?

www.lights.science

University of Basel

Swiss National Science Foundation

KARAKUN

Paperpile

Search

Save search

Search in all documents or in favorites

Sort by Count

1,246 documents found

Sort by Relevance

- Study or data type
- Methodological topic
- Guidance type
- Development process
- Medical context
- Publication date
- Author
- Journal

☆ Developing points-based risk-scoring systems in the presence of competing risks
2016 | Austin, Peter C | Lee, Douglas S | D'Agostino, Ralph B | Fine, Jason P
Our objective is to describe methods for developing simple points-based risk-scoring systems based on the incidence of outcomes in the presence of competing risks....

☆ Timeline cluster: a graphical tool to identify risk of bias in cluster randomised trials
2016 | Caille, Agnès | Kerry, Sally | Tavernier, Elsa | Leyrat, Clémence... Giraudeau, Bruno
This article presents a graphical tool depicting the time sequence and blinding status of the different stages of a cluster randomised trial, together with examples to help researchers describe the st...

☆ Effective Visual Communication for the Quantitative Scientist
2019 | Vandemeulebroecke, Marc | Baillie, Mark | Margolskee, Alison | Magnusson, Baldur
We posit three laws of visual communication for the quantitative scientist. Second, we provide more granular recommendations for good visual display, conve- niently compiled in a single-page reference...

☆ An analysis reveals differences between pragmatic and explanatory diagnostic accuracy studies
2020 | Bossuyt, Patrick M | Olsen, Maria | Hyde, Chris | Cohen, Jérémie F
The objective of this study was to clarify a difference between two approaches while evaluating the diagnostic accuracy ofmedical tests, labeled here as ‘‘pragmatic’’ vs. ‘‘explanatory’’ studies....

☆ Sample Size Requirements for Discrete–Choice Experiments in Healthcare: a Practical Guide
2015 | de Bekker-Grob, Esther W | Donkers, Bas | Jonker, Marcel F | Stolk, Elly A
Discrete-choice experiments (DCEs) have become a commonly used instrument in health economics and patient-preference analysis, addressing a wide range of policy questions. An important question when s...

Select a document to see the details.

sample size calculation

Save search

Search in all documents or in favorites

Clear all

Sort by Count

12 documents found

Sort by Relevance

Study or data type

- randomized controlled trials 12
- cluster-randomized trials 3
- non-specific 3
- cross-over trials 2
- adaptive design trials 1
- clinical practice guidelines 1
- literature reviews 1
- systematic reviews 1

Methodological topic

Guidance type

Development process

Medical context

Publication date

Author

Journal

DELTA2 guidance on choosing the target difference and undertaking and reporting the sample size calculation for a randomised controlled trial

2018 | Cook, Jonathan A | Julious, Steven A | Sones, William | Hampson, Lisa V... Vale, Luke D

Aim to produce updated guidance for researchers and funders on specifying and reporting the target difference ("effect size") in the sample size calculation of a RCT....

A tutorial on sample size calculation for multiple–period cluster randomized parallel, cross–over and stepped–wedge trials using the Shiny CRT Calculator

2020 | Hemming, Karla | Kasza, Jessica | Hooper, Richard | Forbes, Andrew | Taljaard, Monica

"In this article we provide a tutorial on sample size calculation for cluster randomized designs with particular emphasis on designs with multiple periods of measurement and provide a web-based too...

Tutorial in biostatistics: sample sizes for parallel group clinical trials with binary data

2012 | Julious, Steven A | Campbell, Michael J

This article gives an overview of sample size calculations for a single response and a comparison of two responses in a parallel group trial where the outcome is binary....

Sample sizes for clinical trials with normal data

2004 | Julious, Steven A

This article gives an overview of sample size calculations for parallel group and cross-over studies with Normal data...

Choosing the target difference and undertaking and reporting the sample size calculation for a randomised controlled trial – the development of the DELTA2 guidance

2018 | Sones, William | Julious, Steven A | Rothwell, Joanne C | Ramsay, Craig Robert... Cook, Jonathan Alistair

This article reports the development of the DELTA2 guidance on the specification and reporting of the target difference for the primary outcome in a sample size calculation for a RCT....

DELTA2 guidance on choosing the target difference and undertaking and reporting the sample size calculation for a randomised controlled trial

The most common approach is to specify a target difference between the treatments for the primary outcome and then calculate the required sample size. The sample size is chosen to ensure that the trial will have a high probability (adequate statistical power) of detecting a target difference between the treatments should one exist. The sample size has many implications for the conduct and interpretation of the study. Despite the critical role that the target difference has in the design of a RCT, the way in which it is determined has received little attention. In this article, we summarise the key considerations and messages from new guidance for researchers and funders on specifying the target difference, and undertaking and reporting a RCT sample size calculation. This article on choosing the target difference for a randomised controlled trial (RCT) and undertaking and reporting the sample size calculation has been dual published in the BMJ and BMC Trials journals METHODS: The DELTA2 (Difference ELicitation in TriAls) project comprised five major components: systematic literature reviews of recent methodological developments (stage 1) and existing funder guidance (stage 2); a Delphi study (stage 3); a two-day consensus meeting bringing together researchers, funders and patient representatives (stage 4); and the preparation and dissemination of a guidance document (stage 5). RESULTS AND DISCUSSION: The key messages from the DELTA2 guidance on determining the

Key features of LIGHTS

- Open access
- Regular updates
- All methods guidance in one place
- Systematic indexing based on controlled vocabulary
- Methods-specific search filters

Taxonomy and **The**saurus for **H**ealth **Re**search **M**ethodolog**y** (THEREMY)

Applicability

- Generalizability
- External validity
- Transferability
- Context suitability
- Transportability
- Transposability
- Directness
- Indirectness
- …

Subgroup analysis

- Subgroup effect
- Differential effect
- Interaction
- Moderation
- Predictive factor
- Heterogeneity of treatment effects
- …

Next steps

- Machine learning supported search for methods guidance
- Registry for methods guidance under development
- Quality criteria for methods guidance
- Guidance for developers of methods guidance
- Hosting of living methods guidance on www.lights.science


S Schandelmaier


H Ewald


D Lawson


C Schöneberger


G Guyatt


S Chowdhury


J Sebald


J Hirt


M Briel


R Rohner


F Germini


D Papola


S Lin


I Buljan


A Marušić


A Shahabinezhad


L Zeng


J Lobo Lima


H Keibel


J Porzelt


Y Gao