METHODS IN RESEARCH ON RESEARCH

WHAT WE SHOULD CHANGE IN THE WAY WE ARE DOING RESEARCH

MAIN FINDINGS FROM THE H2020 MiRoR PROJECT

MiRoR
Methods in Research on Research

MARIE CURIE ACTIONS

TO DOUG ALTMAN

As a consortium targeting the improvement of methodological quality of clinical research, we have had the honor of working with one of the giants in this field. Doug Altman has left a legacy that will affect the way we are doing research. It is now our responsibility to put in the necessary effort and oversight to follow his path and make research more useful.

Professor Doug Altman was an invaluable member of our consortium for many reasons: not only was he one of the world's leading experts in health research methodology, statistics and reporting, but he was a generous, kind and passionate man. No matter how busy he was, he always found time to listen to ideas, concepts, projects. He was always available to provide insights and suggestions to our students, demonstrating his passion for teaching and mentoring.

Beyond his brilliance, we remember a generous, funny, supportive and down to Earth man.

The MiRoR consortium

The MiRoR project leading to this publication has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 676207

Design and layout: Laura De Nale

Cover image: istockphoto #473916242. Credit: DigitalStorm

Content

Introduction

MIROR is an innovative and ambitious joint doctoral training programme funded by Marie Skłodowska-Curie Actions, dedicated to Methods in Research on Research (MIROR) in the field of clinical research.

“Research on Research”, is an emerging new scientific discipline that aims to reduce waste in research and increase research value. Tens of billions of Euros are wasted each year on studies that are redundant, flawed in their design, never published or poorly reported. The public is the main victim of this waste and reducing waste and increasing value of research represents a major societal challenge.

Our aim was to create, in Europe, an innovative and ambitious multidisciplinary intersectoral joint doctoral training programme, dedicated to Methods in Research on Research (MiRoR) in the field of clinical research. Our proposal involving 15 early-stage researchers, aimed to:

1) prepare students for envisioning the future challenges in clinical research and find innovative solutions to face them,
2) train students to go well beyond the state-of-the-art in their research,
3) help students think differently, taking advantage of the multidisciplinary expertise and intercultural diversity of the network,
4) teach students how to move from research to action and convert knowledge and idea into a product,
5) help students develop skills to match the public and private sector needs and create new professional opportunities

To improve the training of students and confront them with diverse situations and research context and methods, we considered clinical research in its complexity: we tackled various study designs (observational studies, randomised trials, systematic reviews) and various study questions (therapeutic, diagnostic, and prognostic evaluation). In addition, we chose to address different steps of a clinical research projects: **planning, conduct, reporting and peer review**. The research projects of our fellows are grouped into these four areas, each one addressed by a different chapter of this booklet.

The project has been running from March 2016 and it will end on February 2020.

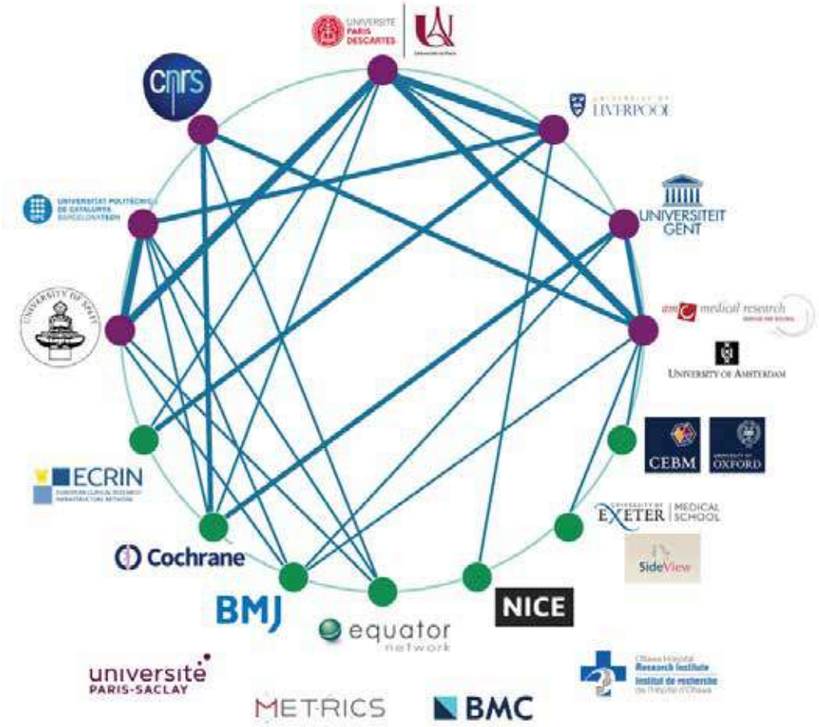
Having most of the MiRoR research fellows concluded their doctoral contracts (some of them are still running but are in the final phase), we think that gathering their findings in one single publication would be an effective way to raise awareness on the main problems existing in the field of clinical research and the proposed solutions resulting from their research projects.

The MiRoR consortium

MIROR brings together **7 world-class research teams** in various disciplines from 6 different European countries, **6 non-academic partners and 6 academic partners** including 3 international research organisations.

Non academic partners are involved in diverse sectors: a large European Clinical Research Infrastructure Network (ECRIN); an international organisation involved in preparing, maintaining and disseminating systematic reviews of the effects of health care (Cochrane); a developer of clinical practice guidelines (the National Institute for Health and Care Excellence); publishers and editors of scientific journals of various sizes and impact factors (British Medical Journal [BMJ] group, BioMed Central Journals [BMC]); a small enterprise providing training and undertaking research (Sideview).

Academic partners involve an international network working to improve the quality of research (the EQUATOR network), the Centre for Evidence-Based Medicine (CEBM) at the University of Oxford, the University of Exeter Medical School, the Université Paris Saclay and the 2 major North American players in this field, the Meta-Research Innovation Center at Stanford University (METRICS) and the Ottawa Hospital Research Institute (OHRI).



This figure describes the MiRoR consortium: beneficiaries are represented in purple, partners in green. The size of the lines between institutions is proportional to the number of secondments.

MiRoR beneficiaries

University Paris Descartes - Université de Paris, France
INSERM, UMR 1153 Epidemiology and Biostatistics Sorbonne Paris Cité Research Center (CRESS), METHODS team
Researchers involved: Isabelle Boutron, Philippe Ravaud, Raphael Porcher, Agnès Dechartres (moved to Sorbonne Université)

The objectives of the METHODS team are to rethink the therapeutic evaluation of chronic diseases, propose new concepts and develop new methodological approaches.
The team explores particularly new methods for developing patients reported outcomes and core outcome set, innovative concepts and methods of precision medicine, new design and analysis methods for routinely collected data, new approaches for evidence synthesis and methods in research on research. We also tackle issues related to multimorbidity, complex interventions.

AMC Medical Research BV (AMR)
University of Amsterdam, The Netherlands
Department of Clinical Epidemiology, Biostatistics & Bioinformatics
Researchers involved: Patrick Bossuyt, Mariska Leeflang, Aeilko Zwinderman

AMR was founded by the Academic Medical Centre of the University of Amsterdam (AMC) to support third-party funded project-based research and to assure the continuity and integrity of the research. The department of Clinical Epidemiology, Biostatistics and Bioinformatics has the mission to conduct research on the etiology, diagnosis, treatment, prognosis, and prevention of disease, to elucidate and understand the underlying molecular mechanisms and to develop methods for these studies.

Universitat Politècnica de Catalunya, Spain
Department of Statistics and Operations Research, UPC Barcelona Tech
Researchers involved: Erik Cobo, Guadalupe Gómez Melis, José Antonio González, Roser Rius

The Department of Statistics and Operations Research at the Universitat Politècnica de Catalunya (UPC) was created in 1987. It has a staff of 40 teachers working at the School of Industrial Engineering of Barcelona (ETSEIB), School of Industrial and Aeronautic Engineering of Terrassa (ETSEIAT), Faculty of Informatics (FIB) and School of Mathematics and Statistics (FME). The department has its own PhD program, and conducts master and graduate courses. The Department is largely involved in research activities, mainly in the field of industrial engineering, the development of intelligent interfaces and automatic data processing, optimization and simulation of flows in networks and applications of the foregoing.

CNRS (Centre national de la recherche scientifique), France
Computer Science Laboratory for Mechanics and Engineering Sciences (LIMSI)
Researchers involved: Aurélie Névéol, Patrick Paroubek

The “Computer Science Laboratory for Mechanics and Engineering Sciences” (LIMSI) is a pluri-disciplinary research laboratory gathering academics and scholars from various scientific fields: primarily from the Engineering and Information Sciences, but also from Cognitive Science and Linguistics. The laboratory was created in 1972 and since then research themes have progressively been expanded to Speech and Image Processing, then to a growing number of themes related to Human-Computer Communication and Interaction on the one hand; to Thermics and Energetics on the other hand.

University of Ghent, Belgium
Department of Applied Mathematics, Computer Science and Statistics
Researchers involved: Els Goetghebeur, Stijn Vansteelandt

Motivated by practical problems, primarily from the bio-medical field, the statistics unit in the department of `Applied Mathematics, Computer Science and Statistics' at Ghent University develops statistical models and studies their statistical and practical properties. We apply existing and new methodology to the design and analysis of important data sets from (bio-medical) researchers, the industry and government. Topics of special interest include causal inference, missing data, noncompliance in clinical trials, statistical genomics and survival analysis. We are further devoted to undergraduate and graduate teaching, and are not only involved in an advanced master in Statistical Data Analysis, but equally in the training of bachelors and masters in the biomedical sciences, biotechnology, informatics, mathematics and pharmacy.

University of Liverpool, UK
Institute of Translational Medicine, Department of Biostatistics
Researchers involved: Paula Williamson, Bridget Young, Catrin Tudur Smith, Carrol Gamble, Jamie Kirkham (moved to University of Manchester)

The Department of Biostatistics carries out methodological and applied research, as well as teaching. Our people are engaged in a broad range of collaborative projects. Main research areas are clinical trials methodology, survival analysis, joint modelling of longitudinal and survival data, statistical pharmacogenetics, multivariate data analysis, meta-analysis, quality of life data analysis, stereology and performance monitoring.

University of Split, Croatia
Department of Psychology School of Humanities and Social Sciences
Researchers involved: Darko Hren

The Department of Psychology at University of Split's Faculty of Humanities and Social Sciences was established in 2013 making it the youngest unit of the Faculty. At the same time, it is the scientifically most productive unit of the Faculty covering research fields in Cognitive, Educational and Social psychology. Members of the Department teach courses in Educational Psychology, Developmental Psychology, Social Psychology, Educational Neuroscience, Psychology of Religion, Statistics, Qualitative Research Methods and Introduction to Scientific Literacy.

Our fellows

Methods for identifying and displaying gaps in health research

Linda Nyanchoka
Université de Paris & University of Liverpool
Supervisor: Raphael Porcher
Co-supervisor: Catrin Tudur Smith
Email: lnyanchoka@gmail.com, Twitter: @LindaNyanchoka

An alternative approach for planning the research of innovative medical tests

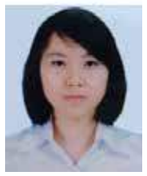
Maria Olsen
University of Amsterdam & Université de Paris
Supervisor: Patrick Bossuyt
Co-supervisor: Els Goetghebeur
Email: m.olsen@amsterdamumc.nl, Twitter: @mariaolsen86

Methods for including participants in core outcome set development

Alice Biggane
University of Liverpool & Université de Paris
Supervisor: Paula Willamson
Co-supervisors: Bridget Young, Philippe Ravaud
Email: alice.biggane@outlook.com, Twitter: @alice_bigbeag

Improving the planning and monitoring of recruitment to clinical trials

Efstathia Gkioni
University of Liverpool & Université de Paris
Supervisor: Carrol Gamble
Co-supervisor: Roser Rius
Email: evigk.biostat@gmail.com, Twitter: @EviGkioni



Impact of mobilising collective intelligence in clinical research planning

Van Thu Nguyen
Université de Paris & University of Liverpool
Supervisor: Isabelle Boutron
Co-supervisor: Bridget Young
Email: nguyenthuvandkh@gmail.com, Twitter: @thuvannguyen1

Improving the use and understanding of causal methods in clinical research

Camila Olarte Parra
University of Ghent and University of Amsterdam
Supervisor: Els Goetghebeur
Co-supervisor: Aeilko Zwinderman
Email: Camila.OlarteParra@ugent.be, Twitter: @colartep

Improving the assessment of Risk of Bias in systematic reviews

Lorenzo Bertizzolo
Université de Paris & University of Amsterdam
Supervisor: Agnès Dechartres
Co-supervisor: Patrick Bossuyt
Email: lorenzo.bertizzolo@gmail.com, Twitter: @l_bertiz

Estimation of causal effects from observational studies: how results obtained from different causal inference methods can be integrated in a meta-analyses approach

Thang Vo Tat
University of Ghent & Université de Paris
Supervisor: Stijn Vansteelandt
Co-supervisor: Raphael Porcher
Email: TatThang.Vo@UGent.be, Twitter: @TatThangVo1

Use of reporting guidelines as an educational intervention for teaching research methods and writing

Melissa Sharp
University of Split & Université de Paris
Supervisor: Darko Hren
Co-supervisor: Guadalupe Gómez Melis
Email: melissaksharp@gmail.com, Twitter: @sharpmelk

Strategies for avoiding “spin” (i.e., distorted reporting) in research reports

Mona Ghannad
University of Amsterdam & Université de Paris
Supervisor: Patrick Bossuyt
Co-supervisor: Isabelle Boutron
Email: mona.ghannad@gmail.com, Twitter: @MonaGhannad

Assisted authoring for avoiding inadequate claims in scientific reporting

Anna Koroleva
Université Paris Saclay & University of Amsterdam
Supervisor: Patrick Paroubek
Co-supervisor: Patrick Bossuyt
Email: aakorolyova@gmail.com, Twitter: @aakorolyova

Text mining for the systematic survey of diagnostic tests in published or unpublished literature

Christopher Norman
Université Paris Saclay & University of Amsterdam
Supervisor: Aurélie Névéol
Co-supervisor: Mariska Leeflang
Email: norman@limsi.fr

Peer review content and communication in biomedical journals

Ketevan Glonti
University of Split & Université de Paris
Supervisor: Darko Hren
Co-supervisor: Erik Cobo
Email: keti.glonti@gmail.com, Twitter: @Keti_2012

Assessing interventions to improve adherence to reporting guidelines in health research

David Blanco
Universitat Politècnica de Catalunya & Université de Paris
Supervisor: Erik Cobo
Co-supervisor: Jamie Kirkham
Email: david.blanco.tena@upc.edu, Twitter: @david_blanco91

Measuring review report quality in health research

Cecilia Superchi
Universitat Politècnica de Catalunya & Université de Paris
Supervisor: José Antonio González
Co-supervisor: Darko Hren
Email: cecilia.superchi@upc.edu, Twitter: @Ceci_Superchi

Innovative «training through action» activities

In addition to the training offered by a typical doctoral program, including courses on advanced research topics and classes on communication, ethics and other soft skills, our fellows have benefitted from cooperative and collaborative learning opportunities, working in small teams to develop their interpersonal skills and reinforce their ability to develop interdisciplinary interactions. In particular, they have participated in the following activities:

Research "speed-dating"

During the first network meeting in Ghent, to favor interaction between students and researchers from other teams and non-academic partners, we organized a 2-hour scientific "speed-dating" session. The consortium was split into 8 groups with a mix of beneficiaries from different teams and partners sitting at each table. Fellows briefly presented their projects at 3 different groups and at each table presentations were followed by a 10-minute Q&A from the panel. This was a unique opportunity for the PhD fellows to interact with all project partners and receive constructive feedback to improve their individual projects.

MiRoR fellows presenting their research projects to the consortium

Common research project

The idea of this activity was to go beyond each fellows' individual research projects and involve the whole group in an innovative project, in a common and transversal effort. Fellows could benefit from this project to enhance their entrepreneurship, professional skills, team building skills and innovative thinking. Fellows were very autonomous in this task: they structured the project, they appointed a coordinating team, they divided the project in different tasks (each led by 2 students), they identified participants in each task and they regularly organized meetings through videoconference. The aim of the project was to investigate the perceptions of and experiences with questionable research practices (QRPs) of PhD students in biomedical disciplines across multiple European countries. An online multinational case-vignette-based survey illustrated by comic strips was developed and it will be disseminated to several European universities in November 2019.

Journal clubs
This activity was led by two fellows from 2 different teams, randomly selected. They had to identify a recently published journal article raising some methodological issues and invite all to comment (online, via videoconference), with the aim of translating the debate into a Letter to the Editor for publication. Aiming to encourage critical thinking and to engage researchers at an early stage to learn how to lead a group and a meeting, this activity was really successful: 7 letters and 3 articles were accepted for publication.

Internal peer review
To train students to their future career as researchers we organised an internal peer-review process: before submitting a manuscript for publication, fellows were invited to send it to another fellow in the consortium who was in charge of writing a peer-review report. This review was sent to the fellow author of the manuscript who had to write an answer to peer-reviewers and modify the manuscript, if needed. This process was very useful because it improved the quality of the manuscript while training students in performing and answering peer review.

Challenge
This activity aimed to initiate entrepreneurial and innovative skills among students. Divided into 4 groups, students worked in teams over 1 year (March 2018 – March 2019) to propose a new intervention that could transform the way research is planned, conducted or reported. The 4 projects were presented to the consortium and to a group of 4 external experts: Sally Hopewell (Centre for Statistics in Medicine \ University of Oxford), Lars Hemkens (Basel Institute for Clinical Epidemiology and Biostatistics, University of Basel), Daniel Strech (Charité Universitätsmedizin Berlin), Casino Gonzalo (Pompeu Fabra University). All fellows received very good feedback from the jury who congratulated everybody for the quality of their work. The group composed of Alice Biggane, Lorenzo Bertizzolo and Ketevan Glonti won the challenge with a project focusing on "Enhancing patient participation in Core Outcome Set development - An alternative method of outcome generation".

The jury and the MiRoR fellows during the Challenge award ceremony

Seminar
The MiRoR consortium and fellows organised a successful seminar dedicated to 'Innovative initiatives to transform research and meet the challenges of the 21st century', taking place at the School of Medicine of the University of Split (Croatia) on October 2, 2018. Guest speakers are leader contributors in the field of meta research: Richard Stephens (Research involvement and Engagement), Ivan Oransky (Retraction Watch), Matthew Westmore (NIHR Evaluation, Trials and Studies Coordinating Centre) and Vicky Hellon (F1000Research). Following the event, the MiRoR fellows could have face-to-face conversations with them on different topics.

RESEARCH ON RESEARCH PLANNING

Planning clinical research is a fundamental step because it affects the relevance, the validity and the feasibility of the research project and, consequently its contribution to society. There is some evidence that the way clinical research is planned is not optimal. The research questions considered are not in line with the needs and priorities of patients, physicians and decision makers.

Our ambition was to rethink research planning and propose innovative solution to improve research value. Particularly, our aim was to:

1. Develop new methods for identifying and displaying gaps in research of therapeutic interventions,
2. Develop alternative approaches for planning the research of innovative medical tests and biomarkers,
3. Explore methods for shifting toward patient-centred research by involving patients in the development of a core outcome set
4. Propose solutions for improving the planning and monitoring of recruitment to clinical trials,
5. Evaluate new ways of planning clinical research through mobilising "collective intelligence" through crowdsourcing.

Each one of these objectives was tackled by a specific research project led by one of the fellows, as presented in the following pages.

Methods for identifying and displaying gaps in health research

Linda Nyanchoka
Université de Paris and University of Liverpool

Background

The current body of research is growing, with over 1 million clinical research articles published from clinical trials alone. Planning a study focusing on the wrong question is however a frequent cause of waste in research. Hence, when planning a health research study, it is crucial to take into account the existing body of research to identify research gaps and prioritize research. Audiences including consumers, patients, researchers, clinicians, advocacy groups, and funders can also benefit from understanding the current status of evidence and research gaps to inform decision and practice. To date there are no standardized methods for systematically identifying or prioritizing research gaps. There is no consensus on what is (or are) the best methodological approach to identify research gaps and prioritize research. The term "research gap" itself can bear different meanings according to the context

Objectives

The aim of this project was to investigate the methods for identifying research gaps in the field of clinical research.

Methods

To develop methodological guidance on methods to identify gaps in health research, we relied on findings from: 1) a scoping review mapping evidence on different definitions reported for the term "research gap" as well as methods used to identify, prioritize and display gaps in health research; 2) a qualitative study aiming to investigate the experience of key stakeholders (researchers, funders, clinicians, clinical guideline developers, public health professionals, commissioners, patients/the public and policymakers) with the definition of research gaps, and practices/ methods used to identify and display such research gaps.

Results

A total of 12 different definitions of a research gap were found, some overlapping, with three cross-cutting themes identified: definitions related to missing information, inadequate information, and insufficient information (1). Studies aiming at identifying gaps were primarily secondary research, and we identified seven specific methods for identifying research gaps. For research prioritization, about half studies used both primary and secondary research, and we identified five specific methods. Finally, we identified 14 methods used to display research gaps. Stakeholder from various background shared with us their conception of gaps in health research and perception on gaps display (2). When there was a common global understanding of the term "gap", conception and expectations differed according to the stakeholder domain of activity. Also, different concepts related to gaps in health research were identified.

Table	Schematic representation	Tree map chart
Box plot		
Bar graph		
Scatter plot		
Funnel plot		
Pie chart		
Word cloud	Geographic map	Mind map
Radial bar plot / polar histogram	Dot plot	Bubble plot/chart

Overview of methods used to display gaps and research priorities

Conclusions

Both studies provide an overview of different methods used to and/or reported on identifying gaps, determining research priorities and displaying both gaps and research priorities. The findings can be adopted to inform the development of methodological guidance on standardizing methods to identify, prioritize and display gaps to inform research and evidence-based decision-making.

1. Nyanchoka L, Tudur-Smith C, Thu VN, Iversen V, Tricco AC, Porcher R. A scoping review describes methods used to identify, prioritize and display gaps in health research. Journal of Clinical Epidemiology. 2019;109:99-110. doi: 10.1016/j.jclinepi.2019.01.005.
2. Nyanchoka L, Tudur-Smith C, Porcher R, Hren D. Key stakeholders' perspectives and experiences with defining, identifying and displaying gaps in health research: a qualitative study protocol. BMJ Open. 2019;9(8):e027926. doi: 10.1136/bmjopen-2018-027926.

An alternative approach for planning the research of innovative medical tests

Maria Olsen
University of Amsterdam and Université de Paris

Background

Many biomarkers are claimed to be useful as innovative medical tests, for e.g. screening and diagnostic applications. Yet, the translational failure of biomarkers indicates that there is a need for improvement. We investigated 3 elements in the evaluation of biomarkers as medical tests. (1) Several authors have pointed out that study design shortcomings may be held responsible for the translational failure, though little empirical evidence of this exist. (2) Assessing a study's validity requires precise communication of study design features. Unfortunately, the definitions of diagnostic test accuracy (DTA) terminology is lacking consensus. (3) Polygenic risk scores (PRS) are now being developed as potential biomarkers, but they need to be evaluated with performance measures that can inform us about their future clinical value.

Objectives

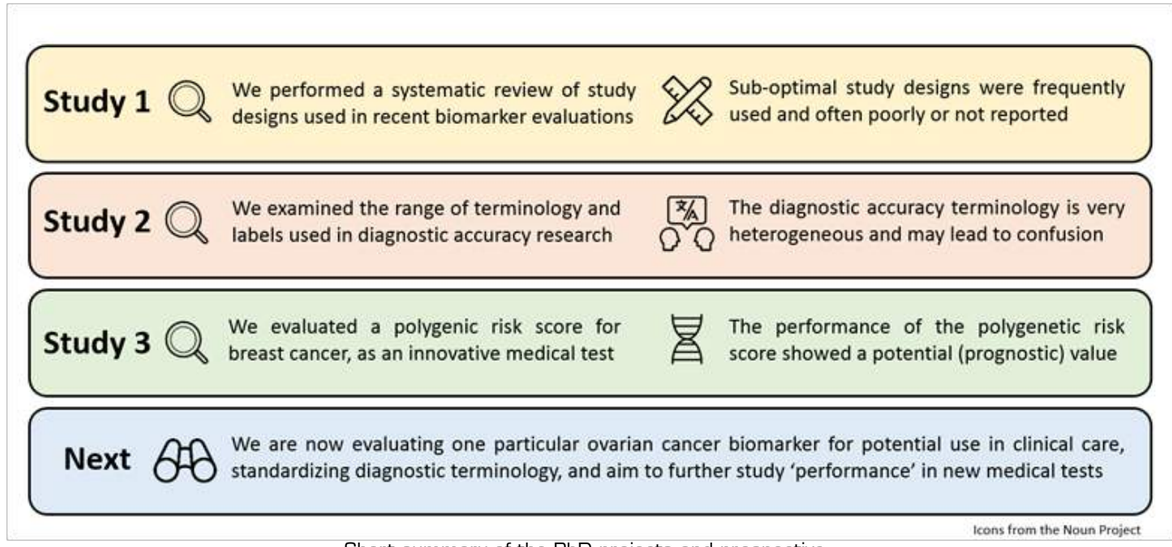
We (1) analysed the study designs that are used in the clinical evaluations of biomarkers, (2) explored the study design terminology in DTA research, and (3) evaluated a breast cancer PRS as a potential biomarker in screening.

Methods

1. We conducted a descriptive systematic review of design features, in recent studies of biomarker evaluations, using ovarian cancer biomarkers as an example. We reported frequencies of study designs, with attention to proposed shortcomings.
2. We abstracted the terminology describing study design features in all the Diagnostic Guidance and corresponding evidence reports from National Institute of Care and Health Excellence (NICE). We assessed the range of terms/labels and also categorized these by design.
3. In collaboration with the Institute of Genomics, Tartu University, we evaluated a breast cancer PRS. We performed survival analyses and assessed the calculated 3 and 5 year absolute risk by measures of discrimination, calibration and re-classification.

Results

1. We analysed 200 articles in our review. Our analysis confirmed that sup-optimal design features were frequently used and that only few studies reported critical information(1).
2. From a total of 17 pairs of NICE guidance and reports, we identified 53 unique design labels, of which 19 (36 %) were specific to DTA designs (2).
3. In the analyses of 30,312 Estonian woman, the PRS-based absolute risk resulted in a close approximation between probabilities and observed proportions, a modest increase in discrimination compared to an age-only model, and an improved classification, for both 3- and 5 year risk (unpublished).



Short summary of the PhD projects and prospective

Conclusions

Our findings confirm the presence of suboptimal design features in evaluations of ovarian cancer biomarkers and the use of heterogeneous and sometimes confusing terminology. We also illustrated that several performance measures can be calculated to inform us about the potential value of a PRS as a biomarker.

If more studies avoided sub-optimal study designs, included adequate performance measures and used standardized terminology, we would improve biomarker evaluations. Correspondingly, we would avoid the introduction of low-quality biomarkers and facilitate the introduction of biomarkers that can improve health care.

1. Olsen M, Ghannad M, Lok C, Bossuyt PM. Shortcomings in the evaluation of biomarkers in ovarian cancer: a systematic review. Clinical Chemistry and Laboratory Medicine. 2019. doi: 10.1515/cclm-2019-0038. (In press)
2. Olsen M, Zhelev Z, Hunt H, Peters JL, Bossuyt P, Hyde C. Use of test accuracy study design labels in NICE's diagnostic guidance. Diagnostic and Prognostic Research. 2019;3:17. doi: 10.1186/s41512-019-0062-9

Methods for including participants in core outcome set development

Alice Biggane
University of Liverpool and Université de Paris

Background

Core outcome sets (COS) are agreed minimum sets of health outcomes that should be measured and reported in all relevant trials. COS development with patient and public input ensures the resulting COS is valid and reliable. Similarly, patient and public input in health outcome selection of clinical guideline development ensures the resulting guidance is relevant to patients.

Objectives

This thesis investigated methods and perspectives surrounding patient and public input in COS and clinical guideline development and identified pointers to support future research in this area.

Methods

A survey of COS developers mapped commonly used methods of patient participation.
A qualitative interview study explored participant experiences of the COS development methods.
An ethnographic study investigated patient and public influence on health outcome selection in clinical guideline development. Discussion with a range of early stage researchers (ESRs) and European consultants enabled reflection on the roles of patients and members of the public in health research.

Results

Survey responses indicated that patient participants were included in 87% (141/162) of published, completed or ongoing COS. The Delphi survey was used singularly or in combination with other methods in 85% (119/140) of projects. The survey findings also highlighted the increasingly global nature of COS development. I interviewed 24 patients and health professionals about their experiences of participation in COS Delphi studies. Some interviewees struggled to understand the purpose of COS and aspects of the Delphi survey. They wanted guidance regarding the use of the scoring system and stakeholder feedback. My ethnography included 230 hours of observations and 18 interviews. This identified the need for continued support and guidance for patients and the public by the committee, specifically, the chairperson, during guideline development. Specific recommendations include the use of plain language, specifically inviting patient and public input, and alternative methods of facilitating involvement including the use of COS previously developed with patient input. Discussion with ESRs and European consultants in combination with the other data in this study identified different perspectives including perceived challenges surrounding the role of patients in methodological health research and health outcome prioritisation.

Pointers

- COS developers should consider the most appropriate medium(s) to communicate their COS Delphi studies information and guidance

  Points to consider: Language used, target audience, health condition

- COS developers need to ensure that the scoring system used is explained in ways that participants can understand.
- COS developers should explain to participants whose perspectives they should consider when scoring in different rounds
- COS developers should explain to participants that in the first round of the Delphi survey they should score outcomes according to their own individual perspective.

  Proxies: In the first round, COS developers should ask proxies to score according to what they anticipate is the perspective of the patient and not from their own perspective as a carer

- COS developers should ask participants in second or subsequent rounds to reflect on the scores of other participants, while also being clear that participants do not have to change their own scores.

  Proxies: should follow the same advice as other participants in second or subsequent rounds

- COS developers can encourage participants to score outcomes they have no experience of to date, but may experience in the future, although an "unable to score" option or equivalent should also be provided for each outcome.
- COS developers should consider the potential influence of their COS Delphi on participants and take appropriate steps to minimise negative effects.
- By understanding what motivates participants into COS Delphi studies, COS developers can devise appropriate recruitment and retention strategies

Conclusions

There has been an increase in patient and public input in COS development, but a lack of parallel increased focus on how to optimise such patient and public input, internationally and across other methodological health research. The findings of this thesis will inform the development of guidance and research in these areas and help to improve methods. International collaboration is also needed to progress patient and public input in health research generally.

1. Biggane AM, Brading L, Ravaud P, Young B, Williamson PR. Survey indicated that core outcome set development is increasingly including patients, being conducted internationally and using Delphi surveys. Trials. 2018;19(1):113. doi:10.1186/s13063-018-2493-y
2. Biggane AM, Williamson PR, Ravaud P, Young B. Participating in core outcome set development via Delphi surveys: Qualitative interviews provide pointers to inform guidance. BMJ Open [Accepted October 2019]
3. Biggane AM, Olsen M, Williamson PR. PPI in research: a reflection from Early Stage Researchers. Research Involvement and Engagement [Accepted October 2019]

Improving the planning and monitoring of recruitment to clinical trials

Efstathia Gkioni
University of Liverpool and Université de Paris

Background

Successfully recruiting the prespecified number of participants in clinical trials remains a difficult challenge. Whilst statistical methods targeting advances in this area have been developed, the application of these methods is limited. In addition, approaches used to predict and monitor recruitment remain frequently unreported. There is a need to identify current practice and bridge the gap between the development of methodology and its implementation.

Objectives

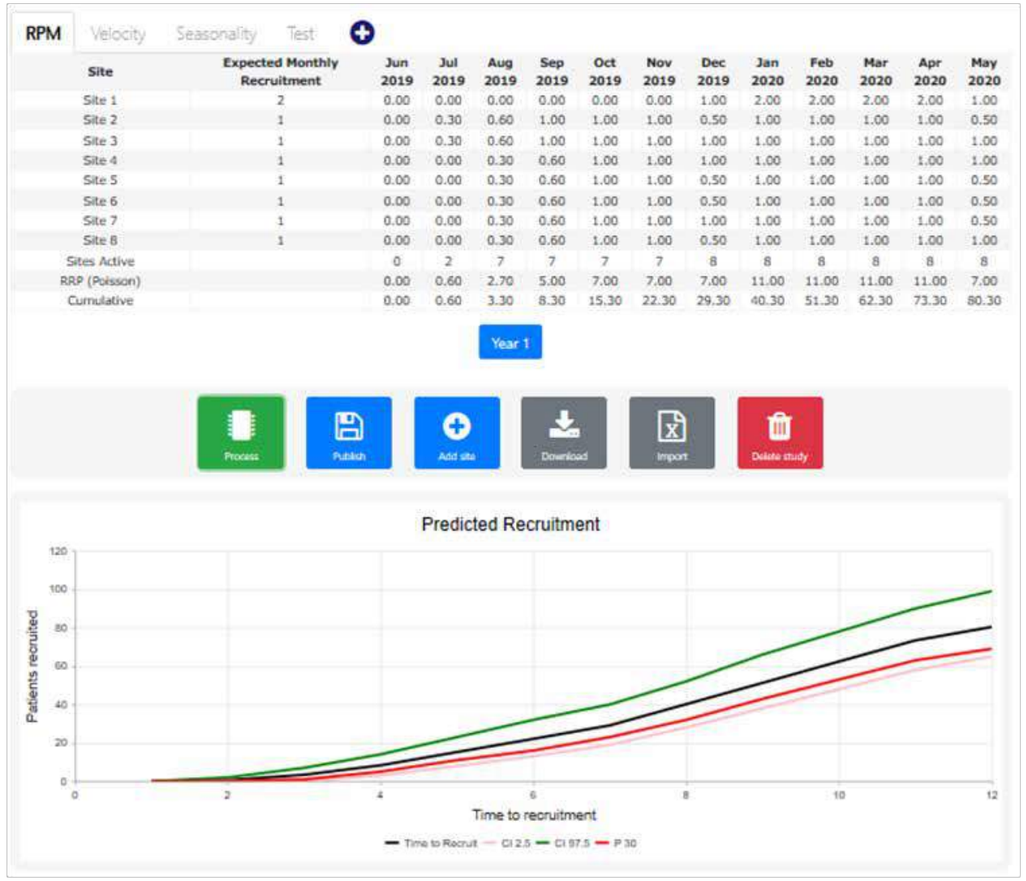
To identify, compare and develop statistical methodology used to predict and monitor recruitment in clinical trials. To develop guidance supported by software with a web-based interface.

Methods

Aiming to identify available methods for recruitment prediction in clinical trials, we conducted two systematic reviews. The first focused on statistical models that can be used at the design stage of the trial and the second on methods used to predict and monitor ongoing recruitment during trial conduct. We conducted a survey of Chief Investigators and a survey of Statisticians across a UK and a European network to identify current practice for recruitment prediction and monitoring. We developed a web-based tool to support recruitment prediction taking into account the needs of research teams and limitations of available models.

Results

In the systematic review of models used for recruitment prediction at the design stage, we assessed and categorised the models according to their nature and ability to incorporate information for recruitment prediction, including time dependent factors such as staggered centre initiations and seasonal variations, and the ability to specify rates per centre or average rates across centres. In the systematic review of methods used during trial conduct we categorised them into methods, comparing expected against actual recruitment including tables and graphs and into statistical models to predict ongoing recruitment based on accrual to date. In the surveys of Chief Investigators and Statisticians we identified that the use of statistical models to predict recruitment was very low (10%). The main reasons are their complexity, the absence of demonstration of their benefits in comparison to simpler approaches used so far and investigators' time pressure. We developed a Shiny application to facilitate the implementation of selected models identified in the literature. To conclude, we also developed a new web-based tool to support recruitment prediction. The tool uses the simple but flexible method adopted by many but also incorporates variation via the Poisson model.



Site	Expected Monthly Recruitment	Jun 2019	Jul 2019	Aug 2019	Sep 2019	Oct 2019	Nov 2019	Dec 2019	Jan 2020	Feb 2020	Mar 2020	Apr 2020	May 2020
Site 1	2	0.00	0.00	0.00	0.00	0.00	0.00	1.00	2.00	2.00	2.00	2.00	1.00
Site 2	1	0.00	0.30	0.60	1.00	1.00	1.00	0.50	1.00	1.00	1.00	1.00	0.50
Site 3	1	0.00	0.30	0.60	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Site 4	1	0.00	0.00	0.30	0.60	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Site 5	1	0.00	0.00	0.30	0.60	1.00	1.00	0.50	1.00	1.00	1.00	1.00	0.50
Site 6	1	0.00	0.00	0.30	0.60	1.00	1.00	0.50	1.00	1.00	1.00	1.00	0.50
Site 7	1	0.00	0.00	0.30	0.60	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.50
Site 8	1	0.00	0.00	0.30	0.60	1.00	1.00	0.50	1.00	1.00	1.00	1.00	1.00
Sites Active		0	2	7	7	7	7	8	8	8	8	8	8
RRP (Poisson)		0.00	0.60	2.70	5.00	7.00	7.00	7.00	11.00	11.00	11.00	11.00	7.00
Cumulative		0.00	0.60	3.30	8.30	15.30	22.30	29.30	40.30	51.30	62.30	73.30	80.30

New recruitment prediction model – A Web-based Tool

Conclusions

Factors affecting patient recruitment may be complex and many; thus, the accuracy of recruitment prediction and the follow up on patient monitoring to ensure that the trial will reach the recruitment target within the time expected, are crucial. Modelling recruitment as a stochastic process allows for uncertainty in the prediction. The new model is based on the Poisson process and by including the stochasticity and uncertainty, it could be proven more accurate that the simplistic approaches used so far. Until researchers implement such models, they are limited in their potential to provide improved predictions.

1. Gkioni E, Rius R, Dodd S, Gamble C. A systematic review describes models for recruitment prediction at the design stage of a clinical trial. Journal of Clinical Epidemiology. 2019;115:141-149. doi: 10.1016/j.jclinepi.2019.07.002.
2. Gkioni E, Dodd S, Rius R, Gamble C. Statistical models to predict recruitment in clinical trials are rarely used by statisticians in UK and European networks. (under review)

Impact of mobilising collective intelligence in clinical research planning

Van Nguyen Thu
Université de Paris and University of Liverpool

Background

Innovative ways of planning and conducting research have emerged recently with promising results. For example, Harvard Medical School organized an ideas competition, which attracted participants from 17 countries who contributed 150 new research ideas for managing type 1 diabetes. These initiatives were based on methods of mobilizing collective intelligence through crowdsourcing.

Objectives

We aimed to explore the methods of mobilising CI and determine how these methods could be used in clinical trial planning.

We defined collective intelligence as shared intelligence emerging when people are mobilized within or outside an organization to work on a specific task that could result in more innovative outcomes than those when individuals work alone

Methods

We first performed a scoping review to describe the methods used to mobilise CI across research disciplines and propose of framework for using these methods. In a second step, to identify barriers and facilitators when in planning and conducting research involving CI, we conducted a large qualitative study using both an online survey and semi-structured interviews of researchers experienced with these new methods. Lastly, we are conducting a proof-of-concept study using methods of mobilising CI to involve patients in trial planning.

Results

The scoping review identified four main reasons of mobilising CI: generate ideas, conduct evaluations, solve problems, and create intellectual outputs. Most projects involved public members who did not necessarily have scientific background. Participants contributed to projects by independent contribution (i.e., no interaction with other participants), collaboration, competitions, and playing games (1). In the qualitative study, researchers from various disciplines highlighted the need of evidence-based guidelines for planning and conducting research mobilising CI. Based on their responses we elaborated practical advice for identifying suitable research problems to be addressed by CI, identifying communities of participants, setting up common rules to mitigate risks of disruptive behaviours (2). The framework in the figure next page outlines steps, highlights risks and challenges of implementing research mobilising CI (1). Using the framework and practical advice, we are conducting a proof-of-concept study in which collective intelligence of patients is mobilised to change the way clinical trials are organised to improve patients' experience of taking part in clinical trials.

Framework of process of mobilizing collective intelligence (CI)

Conclusions

Mobilising CI is an innovative method to increase research efficiency; however, a methodological guidance is lacking. We developed a framework, provided practical advice for implementation and are evaluating its impacts on research planning.

1. Nguyen VT, Benchoufi M, Young B, Ghosn L, Ravaud P, Boutron I. A scoping review provided a framework for new ways of doing research through mobilising collective intelligence, Journal of Clinical Epidemiology, 2019;110:1-11. doi: 10.1016/j.jclinepi.2019.02.007.
2. Nguyen VT, Young B, Ravaud P, Naidoo N, Benchoufi M, Boutron I. Overcoming Barriers to Mobilising Collective Intelligence in Research: Qualitative Study of Researchers with Experience of Collective Intelligence, Journal of Medical Internet Research 2019;21(7):e13792. doi: 10.2196/13792.

RESEARCH ON RESEARCH CONDUCT

How clinical research is conducted is a crucial point. Although the methods used for performing RCTs are well standardised, many important questions remain for observational studies. The main issue lies in overcoming confounding between observed exposure or treatment (often prescribed by indication) and baseline covariates. Over the past decades, important new methods for causal inference have been developed. However, the use and interpretation of these methods can be complex. Similarly important methodological questions are related to meta-analyses of randomised controlled trials, considered the gold standard in evidence-based medicine. Observational studies, particularly those derived from large clinical and administrative databases, should be used routinely in systematic reviews and meta-analysis. Ideally, their information is integrated with that from clinical trials. However, combining the results to arrive at a comprehensive and meaningful estimate of the relevant causal effect and its possible variation over strata (interactions) is a challenge.

WIthin this context our aim was to:

1. Examine accuracy and precision derived from methods currently used in the clinical literature to draw conclusions on causal effects from observational studies; elucidate sources of bias and their likely impact; evaluate how solutions to key problems may follow from alternative state of the art methods at the level of study design and analysis and derive guidance for methods choice.
2. Study reproducibility of the Risk of Bias tool developed by the Cochrane collaboration to assess the most important sources of bias in randomized clinical trials; identify point at weaknesses and suggest improvements for the tool.
3. Develop meta-analysis approaches for randomized trials concerned with risk comparisons, to infer a common, exchangeable effect measure across contributing studies (involving the same `version' of the treatment for the same population) and thereby yielding valid measures of between-study heterogeneity; extend these approaches for observational data and for survival data.

Each one of these objectives was tackled by a specific research project led by one of the fellows, as presented in the following pages.

Improving the use and understanding of causal methods in clinical research

Camila Olarte Parra
University of Ghent and University of Amsterdam

Background

To inform clinical decisions, we need evidence on causal effects of possible interventions. The gold standard for causal effect estimation is a randomised controlled trial (RCT). In many settings, RCTs are not feasible and we must rely on observational studies. The treated and control groups then typically differ in prognostic characteristics. Ignoring this can lead to confounding the treatment effect with the effect of other characteristics. Proper use of causal methods and corresponding language when reporting the findings is critical to avoid misleading conclusions. As a case study, we estimated the survival impact of transplant versus dialysis as initial treatment for end-stage kidney disease.

Objectives

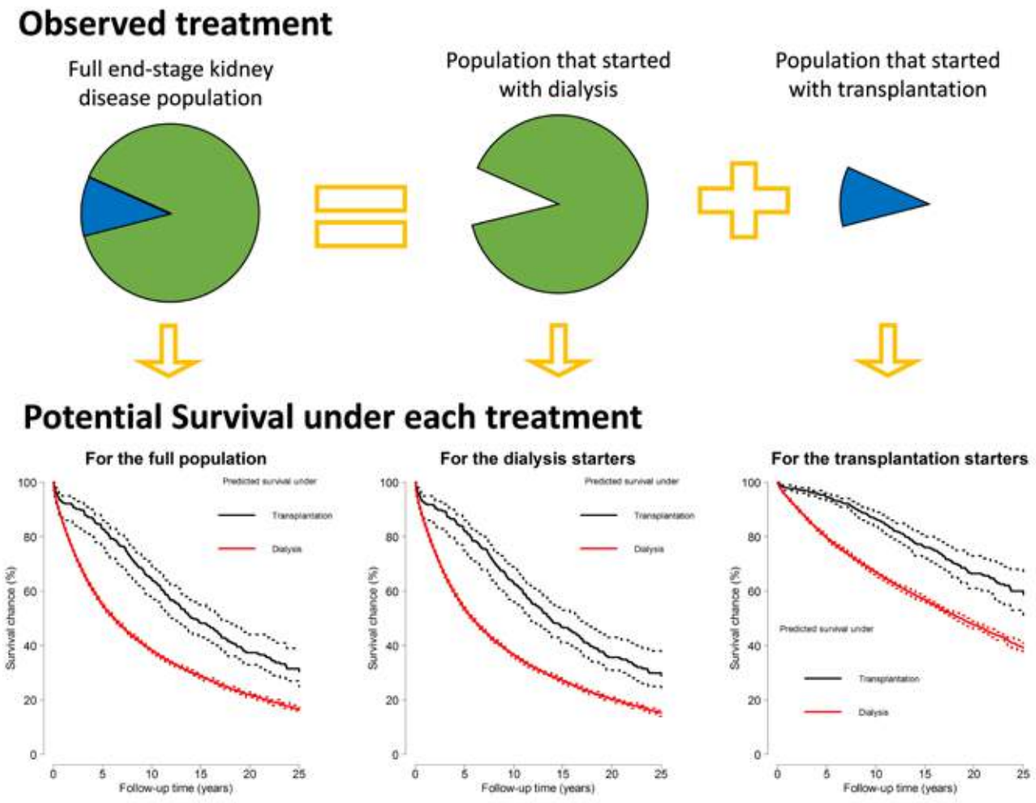
1. To identify methods used to address a clinically relevant causal question and compare their advantages and limitations. 2. To provide guidance for conducting and reporting a causal study.

Methods

The project has 3 parts: systematic review (SR), developing a registry-based study, guidance on reporting causal observational studies. The SR identifies, compares and discusses study designs and statistical methods used for the causal question. From the nationwide Swedish Kidney Registry, we estimate the causal effect of dialysis versus transplant at end stage kidney disease, controlling for the sources of bias identified. Finally, we focus on the reporting of observational studies in a major journal and how the use of causal language (or the lack thereof) could be misleading.

Results

The SR revealed that many studies start follow-up at transplantation and ignore the survival time under dialysis. Others start follow-up at treatment onset but only patients eventually receiving a transplant were eligible. Excluding patients who died on dialysis resulted in immortal time bias.
Our study included patients at the time of end stage renal disease who then started on transplantation or dialysis possibly followed later by transplantation. Assuming we measured the necessary confounders at treatment onset, we quantified the survival advantage of starting with transplantation over the full population and for the starters on either treatment.
By screening abstracts, we identified cases of misleading use of causal language. The main reason was inconsistency between objectives and conclusion, e.g. phrasing the objective as causal but concluding an association or vice versa. In some cases, both lacked causal language but one still suggested to take action given the findings.



Estimating the potential survival under each treatment for different subpopulations.
We used the observed survival under each treatment to predict the potential survival under each treatment for the relevant subpopulations. The survival curves are presented with their respective 95% confidence intervals

Conclusions

Biases in observational studies arise when the start of follow-up does not match the time where treatment is assigned and/or patients become eligible when post treatment events occur. Estimating the survival advantage when receiving kidney transplantation without prior dialysis was possible by matching start of follow-up with treatment onset, i.e. dialysis onset for those without immediate transplant. Eligibility criteria draw on information available at this stage. Starting on dialysis versus immediate transplant is estimated to have a negative impact on survival in the total population with end stage kidney disease varying over the observed treatment groups. There is room to improve causal language when presenting findings from observational studies with a causal aim.

1. Olarte Parra C; Van de Bruaene C; Weynants L; Nagler EV; McAleenan A; Elbers RG; Higgins JP; Goetghebeur E. Pre-emptive versus non pre-emptive kidney transplantationfor end-stage kidney disease, Cochrane Library 2018 https://doi.org/10.1002/14651858.CD013073

Improving the assessment of Risk of Bias in systematic reviews

Lorenzo Bertizzolo
Université de Paris and University of Amsterdam

Background

Assessing the methodological quality of individual studies is a crucial step when conducting a systematic review, because if the individual studies are biased, the SR will be biased. The tool developed by Cochrane (the Risk of bias (RoB) tool) has rapidly become the reference for assessing RoB in randomized trials. Despite its rigorous development and guidelines to use it, previous studies showed that the reproducibility of this tool was suboptimal. The reasons why disagreements in the assessment happened have not been explored in depth.

Objectives

Our project aims to explore the reasons of disagreement in RoB assessment for clinical trials included in two different SRs and highlight potential factors that influence researchers while measuring risk of bias.

Methods

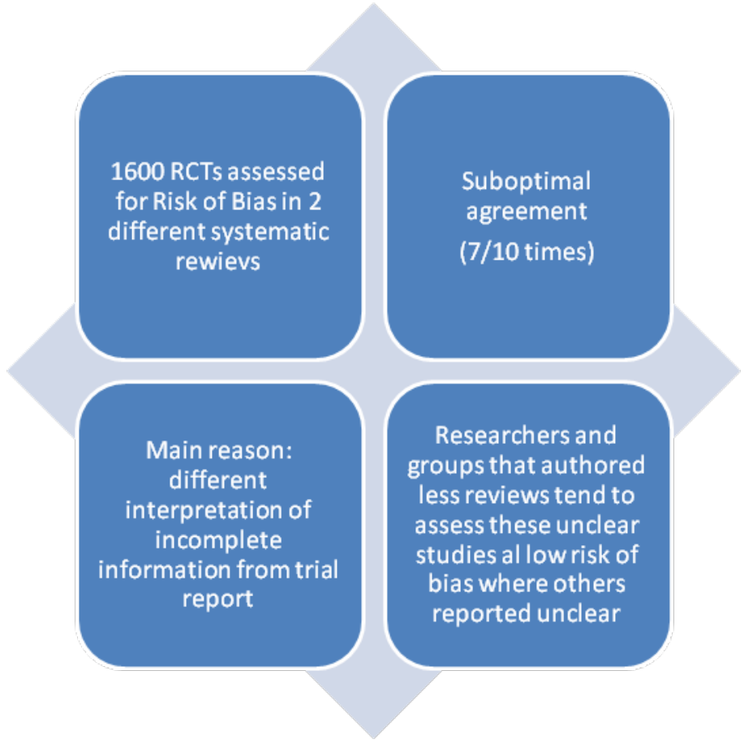
We conducted research-on-research studies using a dataset of recent Cochrane SRs to identify clinical trials that were assessed for RoB in two different SRs.

In our first project, we compared the RoB assessment in the two different SR for the same clinical trial. If there was a disagreement, we evaluated the likely reason, which could access to different or additional information (e.g., contact with trialists) or to a different interpretation of the same information. In case of different interpretation, we evaluated the research report to better understand the reason of disagreement.

The second project focused on the disagreements that were originating from researchers judging an incomplete information from the research report differently. We conducted a matched case–control design to highlight if some factors could have influenced researchers when assessing risk of bias. These factors could be related to the review, its authors or the study itself in comparison with the other studies included in the originating review.

Results

Our results confirm the suboptimal agreement of the RoB tool with researchers agreeing on risk of bias assessment 70% of the times. The most common reason of disagreements was a different interpretation of the same information and these disagreements were most frequently related to incomplete or unclear reporting in the study report. This seemed to push the researchers into using much personal judgement. In the second part of the project, we found that review groups and authors that had completed a lower number of SRs significantly more often assigned a low risk of bias where others reported "unclear". The study year of publication compared to other studies included in the review, was also associated with a different RoB judgment.



Short summary of the project main outcomes

Conclusions

Risk of bias judgements of RCTs included in more than one systematic review differed substantially. Most disagreements were related to a difference in interpretation of an incomplete or unclear description in the study report. A clearer guidance on common causes of incomplete information may be a strategy to improve agreement.

On the other hand, our results indicate that factors related to review and authors may play a role in risk of bias judgements. Specific factors that seem to influence risk of bias judgement are the number of reviews they previously conducted and the number authored by the review group to which they belong. Reviewers also tend to judge a study at unclear risk of bias if they already did so for most of the studies included in the same review. Awareness of these characteristics that could affect risk of bias judgements may help researchers in conducting more reproducible and meaningful assessments of risk of bias.

1. Bertizzolo L, Bossuyt P, Atal I, Ravaud P, Dechartres A. Disagreements in risk of bias assessment for randomised controlled trials included in more than one Cochrane systematic reviews: a research on research study using cross-sectional design. BMJ Open. 2019;9(4):e028382. doi: 10.1136/bmjopen-2018-028382
2. Bertizzolo L, Bossuyt P, Atal I, Ravaud P, Dechartres A. With an incomplete or unclear report, studies can be classified at high or low risk of bias: a research on research study. (submitted)

Estimation of causal effects from observational studies: how results obtained from different causal inference methods can be integrated in a meta-analyses approach

Vo Tat Thang
University of Ghent and Université de Paris

Background

Meta-analysis is a cornerstone of comparative effectiveness research, as it allows synthesizing the evidence from multiple randomized controlled trials and inferring the effect of interventions with increased precision. A key issue in meta-analysis is heterogeneity, which arises due to the fact that studies included in a systematic review often differ to some degree in the case-mix of participants, the variant of the intervention, settings and outcome. Among these factors, case-mix difference between trials can be quite a nuisance as it can make the result from different trials difficult to pool.

Objectives

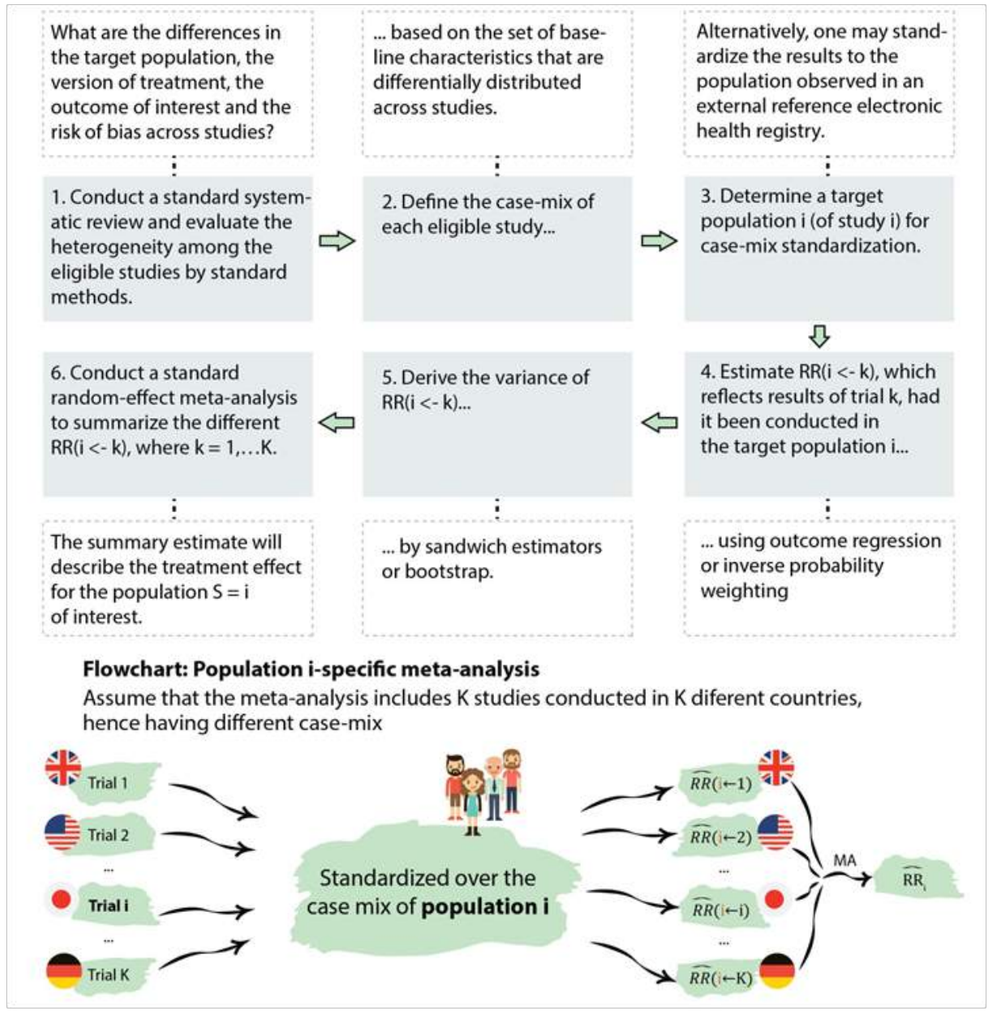
We propose a new approach for individual patient data (IPD) meta-analysis of randomized controlled trials that allows a) to control for differences in the case-mix across studies and reduce heterogeneity, and b) to infer the treatment effect for a population that is well-defined in terms of case mix.

Methods

The above objectives are achieved by standardizing the results from the different trials to the same patient population, e.g. the patient population observed in one of the trials or any other population of interest, based on direct standardization using either outcome regression (OCR) or inverse probability weighting (IPW) [1]. The standardized results are then meta-analyzed as in a classical two-stage IPD meta-analysis. We illustrate the new approach by conducting a meta-analysis of numerically simulated RCTs that evaluate a binary treatment versus control with respect to a binary outcome. The new framework is then applied to reanalyze a published IPD meta-analysis evaluating the effect of vitamin D on the risk of respiratory infection.

Results

Both the OCR-based and IPW-based estimators are effective for case-mix standardization across different populations [1]. As an added advantage, the proposed approaches enable one to decompose the overall heterogeneity between the trial results into two different sources, which the usual approaches to IPD meta-analysis do not provide: case-mix heterogeneity (i.e. arising when the treatment effect is modified by one or more of the factors used to define case-mix) and 'beyond case-mix heterogeneity' (i.e. arising due to the difference between studies in design or methodological aspects) [1].



Conclusions

Assessing the impact of case-mix variation across the eligible studies is an important task in every meta-analysis. This can be done via standardizing evidences across different trials to one well-defined population (in terms of case-mix) before summarizing them. The proposed approaches potentially improve the quality of systematic reviews and meta-analyses in practice, and should be further investigated in future works.

1. Vo TT, Porcher R, Chaimani A, Vansteelandt S. A novel approach for identifying and addressing case-mix heterogeneity in individual patient data meta-analysis. Research Synthesis Methods 2019 (In press). doi: 10.1002/jrsm.1382
2. Vo TT, Superchi C, Boutron I, Vansteelandt S. The conduct and reporting of mediation analysis in recently published randomized controlled trials: results from a methodological systematic review. Journal of Clinical Epidemiology. 2019;117:78-88. doi: 10.1016/j.jclinepi.2019.10.001

RESEARCH ON RESEARCH REPORTING

Inadequate reporting is a frequent cause of research waste. Serious problems include but are not limited to non-reporting or delayed reporting of whole studies, uncertainty about research methods, selective reporting of outcomes, inadequate reporting of harms, omissions from or misinterpretation of results in abstracts, and the use of "spin" (i.e. distorted reporting and interpretation of the study results to convince the reader of the beneficial effect of the treatment not supported by the main findings).

Over the last decade, "reporting guidelines" have been developed to improve the transparency and quality of reporting the results of clinical trials, observational studies, etc. These guidelines were published simultaneously in several leading medical journals and have received support from the World Association of Medical Editors, the Council of Science Editors, and the International Committee of Medical Journal Editors (ICMJE), and several journal editors. Paradoxically, the adherence of authors to these guidelines is low and the quality of reporting remains insufficient.

Within this context our aim was to:

1. Use reporting guidelines as an educational intervention for teaching research methods and writing;
2. Identify, develop and evaluate strategies for preventing, eliminating, or limiting «spin» (i.e., distorted reporting) in research reports;
3. Develop an automated tool aimed at identifying and avoiding spin in medical literature; and
4. Construct methods to automate, semi-automate or assist the screening process in systematic reviews of diagnostic test accuracy studies.

Each one of these objectives was tackled by a specific research project led by one of the fellows, as presented in the following pages.

Use of reporting guidelines as an educational intervention for teaching research methods and writing

Melissa Sharp
University of Split and Université de Paris

Background

Poor reporting of biomedical research has been a persistent problem. In many fields, results fail to be reproduced and replicated, often due to incomplete reporting. Reporting guidelines were created to establish minimum criteria to report study results. In 2007, the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement was created for observational studies. It contains 22-items and guidance for authors of case-control, cohort, and cross-sectional studies. It is supported by many journals, and has been expanded upon for specific fields and methodologies through the creation of "extensions», but lack of awareness is widespread and it is unclear how, why and when authors they use STROBE. Furthermore, its potential to serve as a basis for an educational intervention is unexplored.

Objectives

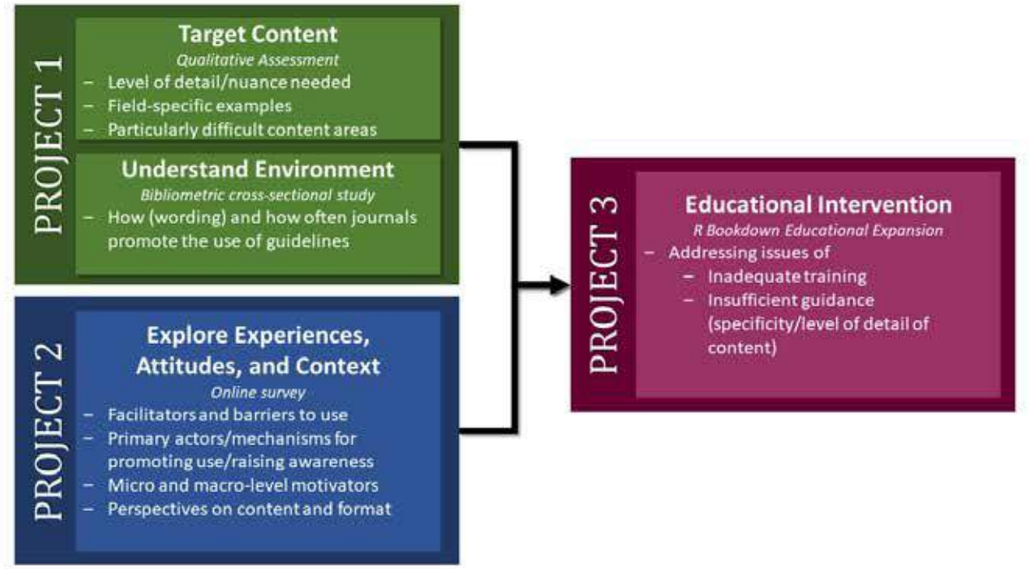
To investigate the evidence, contextual and facilitating factors necessary for the development of a structured educational intervention based on the STROBE statement for teaching observational research methods and reporting.

Methods

To transform STROBE from a reporting guideline into an educational tool and implement research successfully, the Promoting Action on Research Implementation in Health Services (PARIHS) knowledge translation strategy was used to guide evaluations of the evidence, context, and facilitators. First, a qualitative assessment was performed on content in the STROBE extensions to identify strengths and weaknesses in checklist items. Concurrently, a cross-sectional bibliometric study determined endorsement rates of extensions to establish the publishing context that authors are working in. Next, an online survey was distributed to assess researcher's awareness of, experiences with, and attitudes towards STROBE. This established the facilitators, timing and motivators (context), and perceptions (evidence) of use.

Results

Content in the STROBE extensions is sometimes redundant – potentially indicating a poor understanding of certain concepts. The extensions are endorsed at low rates. Journals are largely not endorsing STROBE and the language that they use is vague. The online survey found a large disagreement regarding the level of specificity desired in STROBE and its usefulness. Generally, authors were not opposed to using it but there were often no strong motivating force. Their coauthors did not use it and many journals are not requiring it. Authors also held some internal views that are detrimental to the promotion of STROBE, such as the excessive self-confidence.



Short summary of the PhD projects and main objectives

Conclusions

Work from the first two projects provided the content and support for an educational intervention that will be integrated within the writing process, accessible by a worldwide audience, and open-source and editable. It is built using R Markdown and Github and is available online.

1. Sharp MK, Tokalić R, Gómez G, Wager E, Altman DG, Hren D. A cross-sectional bibliometric study showed suboptimal journal endorsement rates of STROBE and its extensions. Journal of Clinical Epidemiology 2019;107:42-50. doi: 10.1016/j.jclinepi.2018.11.006
2. Sharp MK, Bertizzolo L, Rius R, Wager E, Gómez G, Hren D. Using the STROBE statement: survey findings emphasized the role of journals in enforcing reporting guidelines. Journal of Clinical Epidemiology. 2019;116:26–35. doi:10.1016/j.jclinepi.2019.07.019

Strategies for avoiding "spin" (i.e., distorted reporting) in research reports

Mona Ghannad
University of Amsterdam and Université de Paris

Background

An essential step in the scientific process is publication in peer-reviewed journals. Ethically, research findings should be disseminated completely and accurately. However, authors may intentionally or non-intentionally "spin" their results for readers. "Spin" refers to reporting practices that distort the presentation or interpretation of results. A consequence of biased representation of results in scientific reports is that the published literature may suggest stronger evidence than is justified.

Objectives

This project aims to identify, develop and evaluate strategies for preventing, eliminating, or limiting "spin" in clinical research.

Methods

1. We performed a systematic review to document the prevalence of spin in recent evaluations of the clinical performance of biomarkers in ovarian cancer. The primary aim of our study was to evaluate the presence of spin, further categorized as misrepresentation and overinterpretation of study findings, in the first 200 studies reporting the performance of the discovered biomarker. In addition, we also evaluated facilitators of spin (i.e., practices that would facilitate overinterpretation of results).
2. We are conducting a two-arm randomized controlled trial to evaluate the effectiveness of an intervention for reducing spin in the abstract conclusion of primary research manuscripts submitted to BMJ Open. In the intervention group, the authors will receive additional instructions with the peer reviewers' comments, inviting them to check for and remove spin in the abstract of their revised manuscript. Our primary outcome is spin in the abstract conclusion of the revised manuscript.

Results

In our 200 analyzed studies, 140 (70%) contained one or more forms of spin in the title, abstract or main text conclusion, exaggerating the performance of the biomarker; 75 (38%) had two or more forms of spin, and only sixty studies (30%) had no form of spin in the article, based on our criteria.

In terms of facilitators of spin, we observed that none of the 200 analyzed studies reported a sample size justification or discussed any potential harms, and most of the articles did not prespecify a positivity threshold for continuous biomarkers.

Forms of Spin	Spin frequency, n=200 n (%) [95% CI]
Misinterpretation	
Other purposes of biomarker claimed not pre-specified and/or investigated	65 (32.5%[26%- 40%])
Mismatch between intended aim and abstract or main text conclusion	57 (28.5%[23%- 35%])
Other benefits of biomarker claimed not pre-specified and/or investigated	10 (5%[3%- 9%])
Extrapolation from study participants to a larger or a different population	10 (5%[3%- 9%])
Misrepresentation	
Incorrect presentation of results in the abstract or main text conclusion	40 (20%[15%- 26%])
Mismatch between results reported in abstract and main text	33 (16.5%[12%- 23%])
Mismatch between results reported and the title	11 (5.5%[3%- 10%])

Table 1: Actual forms of spin in clinical studies evaluating performance of biomarkers in ovarian cancer

Potential Facilitators of Spin	Spin frequency, n=200 n (%) [95% CI]
Not stating sample size calculations	200 (100% [98%-100%])
Not mentioning potential harms	200 (100% [98%-100%])
No pre-specified threshold of continuous biomarker	84/164* (51.2% [43% - 59%])
No imprecision or statistical test reported	26 (13% [9%-19%])
Study objective not reported or unclear	24 (12% [8%-18%])

Table 2: Facilitators of spin in clinical studies evaluating performance of biomarkers in ovarian cancer

Conclusions

Our review indicates that spin or biased reporting and interpretation is prevalent in recent clinical evaluations of biomarkers in ovarian cancer. These results indicate a need for strategies to minimize biased reporting and interpretation, which we are investigating in our ongoing intervention study. Efforts to prevent or reduce biased and incomplete reporting in biomedical research should be undertaken with vigor and in unison, given the intricate complexities that involve multiple players. Researchers and authors, peer reviewers, and journal editors unboundedly share responsibility.

1. Ghannad M, Olsen M, Boutron I, Bossuyt PM. A systematic review finds that spin or interpretation bias is abundant in evaluations of ovarian cancer biomarkers. Journal of Clinical Epidemiology. 2019;116:9-17. doi: 10.1016/j.jclinepi.2019.07.011
2. Ghannad M, Yang B, Leeflang M, Aldcroft A, Bossuyt PM, Schroter S, Boutron I. Evaluating an editorial intervention to reduce spin in the abstract conclusion of manuscripts: a randomized controlled trial. Retrieved from osf.io/y2ewa (2019, August 6)

Assisted authoring for avoiding inadequate claims in scientific reporting

Anna Koroleva
Université Paris Saclay and University of Amsterdam

Background

Randomized controlled trials (RCTs) are a type of clinical trials that study a new (experimental) intervention by comparing it to a standard intervention. RCTs are believed to be a very robust trial design providing high-quality evidence for health care. However, reporting of findings of RCTs can be distorted by presence of spin, i.e. presenting the experimental intervention to be safer or more effective than the experiments justified. Spin in RCTs can have negative impact on clinical practice, as it was shown to make clinicians overestimate the intervention. Besides, spin in research articles can result in spin in press releases and health news, raising false hopes and expectations among the general public. In 2016-2019, spin was shown to be present in RCTs with non-significant primary outcome in domains such as surgery (40%), cancer (47%), obesity (46.7%), otolaryngology (70%), anaesthesiology (32,2%) and wound care (71%).

Objectives

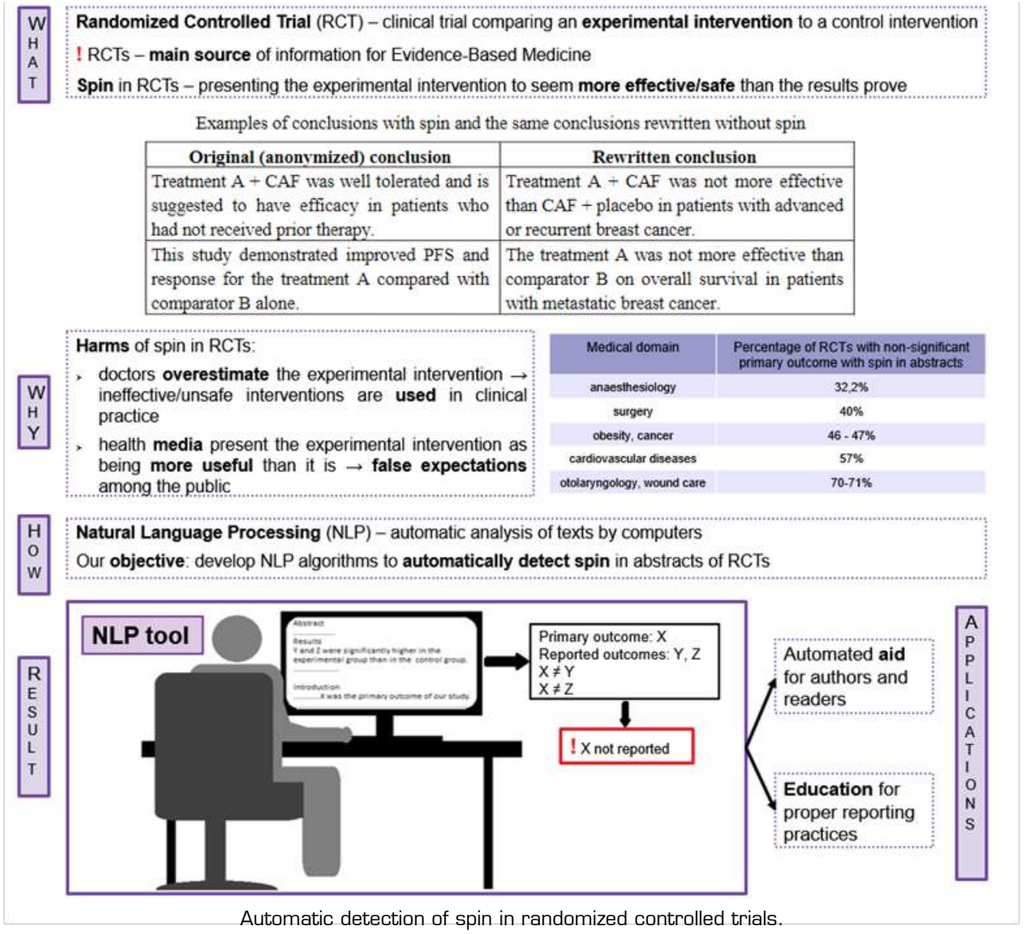
The high prevalence of spin demonstrates that it often remains unnoticed by journal editors and peer reviewers. Authors of articles can themselves be unaware of spin, introducing it unintentionally. Thus, our project was aimed at developing algorithms that can be used as automated assistance for authors and readers of scientific articles in the task of spin detection.

Methods

As spin can be viewed as a textual phenomenon, we used Natural Language Processing (NLP) methods to analyse texts of articles reporting RCTs. We conducted a linguistic study of textual expressions related to various types of spin and identified text elements that need to be identified to detect spin. We combined rule-based and machine learning methods to extract spin-related information. For the core tasks of our pipeline, we employed a deep learning method that consists in fine-tuning pre-trained language models on task-specific annotated data. We manually annotated the corpora for the key tasks.

Results

We developed a set of algorithms for identifying potential spin and related information. The algorithms include text structure analysis (identification of abstracts), extraction of entities (trial outcomes, significance levels), relations between the entities, sentence classification (detection of sections within abstracts, detection of specific types of statements related to spin). Our algorithms achieved operational performance for detecting relevant phenomena (F-measure from 79 to 98%). We developed a simple interface that allows to run the algorithms, visualize their outputs and generate a report.



Automatic detection of spin in randomized controlled trials.

Conclusions

The developed tool can be used by authors and peer reviewers of scientific articles as assistance in spin detection, thus helping to improve the quality of research results reporting. The tool and the annotated datasets are freely available.

1. Koroleva A, Paroubek P. Extracting relations between outcomes and significance levels in Randomized Controlled Trials (RCTs) publications. Proceedings of the 18th BioNLP Workshop and Shared Task, 2019, doi:10.18653/v1/W19-5038
2. Koroleva A, Kamath S, Paroubek P. Measuring semantic similarity of clinical trial outcomes using deep pre-trained language representations. Journal of Biomedical Informatics: X, 2019;4:100058. doi:10.1016/j.yjbinx.2019.100058

Text mining for the systematic survey of diagnostic tests in published or unpublished literature

Christopher Norman
Université Paris Saclay and University of Amsterdam

Background

The demand and production of systematic reviews is increasing rapidly. PubMed indexed 17,254 new systematic reviews in 2018 alone, and this number has increased more than five-fold since 2009. While the demand for systematic reviews is growing, the number of publications that systematic reviews need to sift through is also increasing at a similarly break-neck pace. We today spend more time and money producing new systematic reviews than we ever have.

Authors conducting systematic reviews face issues throughout the systematic review process. It is difficult and time-consuming to search and retrieve, collect data, write manuscripts, and perform statistical analyses.

Objectives

In this project we have attempted to explore methods for performing systematic reviews quicker, cheaper, and more efficiently. At the same time, systematic reviews still require a thorough, objective, and reproducible methodology to avoid bias.

While we are attempting to make the process more expedient, we are also striving to uphold the same methodological rigor of the process.

Methods

In this project we have examined how machine learning methods can be used to reduce this workload, how such methods can be made to work, and how they can fit into different systematic review contexts and settings. During this project we have investigated three different models, for slightly different systematic review contexts:

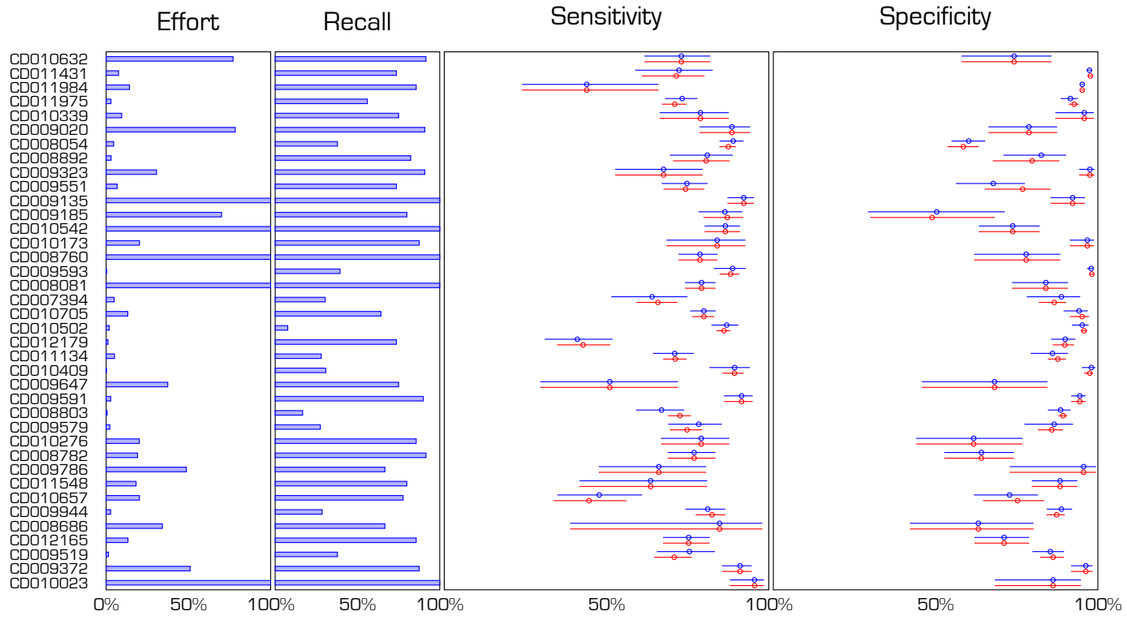
1) A static model is trained on the inclusion/exclusion decisions of references screened in previous systematic reviews.

2) An active model using active learning to improve its performance throughout screening; and

3) A stacked model combining the static and active models to achieve the best of both.

Results

We have presented a screening automation system that can be used in a variety of systematic review contexts - ranging from review updates to reviews conducted de novo. The system is general in purpose, and performs well on reference screening datasets on clinical NLP, drug class efficacy (intervention studies), DTA studies, and core outcome set development. The system is furthermore highly customizable, and the underlying preprocessing pipeline and classification or ranking algorithms can be changed to fine-tune the system for specific systematic review topics or contexts.



The impact of screening prioritization and stopping criteria on meta-analyses.
Difference in meta-analysis results for the largest meta-analysis in each systematic review using a combination of stopping criteria (Displacement (0.01) OR Relevant (n = 15) OR Found/Effort (1/1,000)).
Ten systematic reviews did not include any meta-analysis based on three or more studies (in PubMed) and were therefore excluded from the results.
Effort denotes the fraction of candidate references screened. Recall denotes the fraction of identified relevant studies.
Blue data points correspond to the simulated results using early stopping. Red data points correspond to results without early stopping, i.e. equivalent to current practice (which would have 100% effort and 100% recall).

Conclusions

Systematic review automation methods can be used in systematic reviews without fundamentally altering the process. Screening reduction method can be used as an extra search filter, leaving the remainder of the review process identical to the conventional process, including screening in random order, and the use of standard reference managers like EndNote.

The accuracy of the screening process, and the impact it has on the results and conclusions of the review can be measured prospectively through the screening process using cumulative meta-analyses. This requires modifying the systematic review process to perform data extraction and meta-analyses concurrently, but can lead to substantial improvements over traditional stopping criteria for screening automation.

1. Norman CR, Gargon E, Leeflang MMG, Névéol A., Williamson PR. Evaluation of an automatic article selection method for timelier updates of the Comet core outcome set database. Database. Oxford University Press.
2. Norman CR, Leeflang MMG, Porcher R, Névéol, A. Measuring the impact of screening automation on meta-analyses of diagnostic test accuracy. Systematic Reviews. 2019;8(1):243. doi: 10.1186/s13643-019-1162-x.

RESEARCH ON PEER REVIEW

Peer review is the cornerstone of the scientific evaluation of manuscripts submitted for publication. It aims to help editors publish scientifically sound articles but also to improve their quality and reporting. Peer review is amongst the oldest methods used in medical science and has been one of the least studied until recently. However, the effectiveness and sustainability of the peer review system has been challenged. For example, peer reviewers often fail to detect important deficiencies in the reporting of the methods and results of RCTs.

Several strategies have been proposed to improve peer reviewers' performance such as training peer reviewers or change in the review process (e.g. revealing of peer reviewers' identities). Nevertheless, we have little evidence about the efficiency of these measures and interventions. Although some isolated authors used their own outcome to measure the quality of the review, a validated and consensed measure of the quality of the review is still needed.

Our aim was to explore the peer review and editorial process and particularly to:

1. Analyze the process of communication between the authors, editors and peer reviewers by identifying the roles and tasks of peer reviewers in the manuscript review process in biomedical journals and exploring the journal editors' understanding of the roles and tasks of peer reviewers;
2. Investigate and assess what actions can be taken to improve the adherence to reporting guidelines of studies in health research.
3. Explore and develop a new tool for measuring the quality of a peer review report.

Each one of these objectives was tackled by a specific research project led by one of the fellows, as presented in the following pages.

Peer review content and communication in biomedical journals

Ketevan Glonti
University of Split and Université de Paris

Background

Published findings may have a direct impact on clinical practice and inform policy, so it is crucial that the peer review process in biomedical journals works well. Peer reviewers play a key role in the manuscript review process, providing expert knowledge that informs the decision of the journal editor. However, their roles and tasks are often poorly defined and communicated. Studies suggest that the most important tasks in peer review, as perceived by peer reviewers, are not congruent with the tasks most often requested by journal editors. Clarity around the expected content and appropriate communication between key stakeholders is important as it may influence the quality of peer reviewer reports.

Objectives

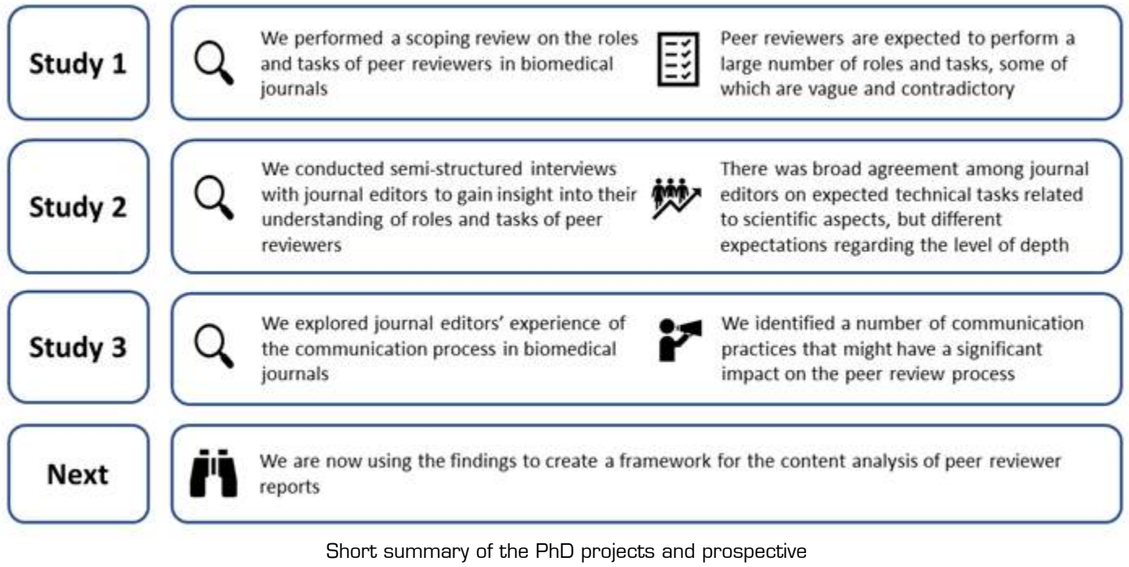
This study aimed to provide clarity around the expected content and communication practices within the editorial peer review process in biomedical journals to identify new ways of improving peer review.

Methods

We conducted a scoping review of the literature to describe the roles and tasks of peer reviewers. We also conducted a qualitative study using semi-structured interviews to gain insight into journal editors' understanding of the roles and tasks of peer reviewers, and to explore their experience of the communication process. Lastly, we used the findings to create a framework for the content analysis of peer reviewer reports.

Results

We analysed 209 articles in our scoping review, which confirmed the lack of clarity of peer reviewers' roles and tasks and highlighted incongruities between the respective positions of the peer reviewer and journal editor. These findings were corroborated through the insight gained from interviewing 56 biomedical journal editors. Journal editors' understanding of the roles and of tasks of peer reviewers is profoundly shaped by each journal's unique context and characteristics, including financial and human resources and journal reputation or prestige. There was a broad agreement on expected technical tasks related to scientific aspects, but there were different expectations with regards to the level of depth and detail peer reviewers should provide. The absence of effective communication is apparent in the poor transfer of critical information and a number of missed opportunities to improve the quality of peer reviewer reports were evident. We also illustrate missed opportunities for journal editors to engage with peer reviewers to clarify the expected roles and tasks, and identified a number of communication practices that might significantly impact the peer review process.



Short summary of the PhD projects and prospective

Conclusions

These findings demonstrate that peer review is a complex, social process. If peer review is to improve, it is necessary to move away from simple interventions, and instead implement changes that can affect multiple components of the peer review system simultaneously.

1. Glonti K, Cauchi D, Cobo E, Boutron I, Moher D, Hren D. A scoping review on the roles and tasks of peer reviewers in the manuscript review process in biomedical journals. BMC Medicine. 2019;17(1):118. doi.org/10.1186/s12916-019-1347-0
2. Glonti K, Boutron I, Moher D, Hren D. Journal editors' perspectives on the roles and tasks of peer reviewers in biomedical journals: qualitative study. BMJ Open 2019;0:e033421. doi:10.1136/bmjopen-2019-033421

Assessing interventions to improve adherence to reporting guidelines in health research

David Blanco
Polytechnic University of Catalonia and Université de Paris

Background

The lack of transparency and accuracy of research reports is one of the main factors causing waste in research. Reporting guidelines (RGs) are sets of recommendations for authors on how to report research methods and findings in a way that no relevant information is missing. So far, biomedical authors' adherence to RGs is inadequate.

Objectives

1. To identify, classify, and analyse interventions to improve adherence to RGs, and to determine the existing gaps in research on the evaluation of interventions.
2. To explore biomedical editors' perceptions of different interventions to improve adherence to RGs that have been or can be implemented at various points in the editorial process.
3. To evaluate the impact of an editorial intervention in a real context in collaboration with BMJ Publishing Group.

Methods

1. Scoping review of the published and grey literature. Development of a typology of interventions that can be performed at different stages of research.
2. Survey for biomedical journal's editors that explored (i) the current practice of their journals, (ii) their perceptions of the ease of implementation and the potential effectiveness of different interventions, (iii) the barriers and facilitators associated with these interventions, and (iv) suggestions for future interventions and incentives.
3. Randomised controlled trial (RCT) in collaboration with BMJ Open. The intervention consisted of evaluating the consistency between the submitted CONSORT checklist and the manuscript, and to provide feedback to authors. The control group underwent usual peer review.

Results

1. 31 interventions grouped into five categories: (A) training on the use of RGs, (B) improving understanding, (C) encouraging adherence, (D) checking adherence and providing feedback, and (E) involvement of experts. Research gaps (1) on training on the use of RGs and improving their understanding, (2) at early stages of research and (3) after the final acceptance of the manuscript.
2. Involving trained editors or administrative staff was deemed the potentially most effective intervention. However, it was considered moderately difficult to implement due to logistic and resource issues. Participants believed that checking adherence to RGs goes beyond the role of peer reviewers and could decrease the overall quality of reviews.
3. 24 RCTs were included (12 intervention, 12 control). The estimated effect (on a 0-10 completeness of reporting scale) of performing the intervention compared to the usual peer review process was 1.79 (95% CI: 0.39-3.18).

Outcome	Intervention Mean (SD)	Control Mean (SD)	Mean difference* (95% CI)
Completeness of reporting (0-10 scale) with imputation (n=24)	8.75 (1.85)	7.09 (1.79)	1.79 (0.39-3.18)
Completeness of reporting (0-10 scale) without imputation (complete case analysis, n=18)	9.31 (1.27)	7.34 (1.82)	2.19 (1.00-3.44)

*Adjusted by baseline score.

Summary of Project (3) results

Conclusions

Biomedical journals need to involve trained editors or administrative staff in the process of ensuring adherence to RGs. Our RCT showed that this may significantly increase the transparency and accuracy of published research.

Moreover, further evaluations of interventions by different stakeholders (research funders, ethics boards, universities, or journals) at different research stages are needed. These evaluations could take into account the points raised in our scoping review and our survey.

1. Blanco D, Altman D, Moher D, Boutron I, Kirkham JJ, Cobo E. Scoping review on interventions to improve adherence to reporting guidelines in health research. BMJ Open. 2019 May 9;9(5):e026589. doi: 10.1136/bmjopen-2018-026589.
2. Blanco D, Hren D, Kirkham JJ et al. A survey exploring biomedical editors' perceptions of editorial interventions to improve adherence to reporting guidelines [version 1; peer review: awaiting peer review]. F1000Research 2019, 8:1682. https://doi.org/10.12688/f1000research.20556.1

Measuring review report quality in health research

Cecilia Superchi
Polytechnic University of Catalonia and Université de Paris

Background

Editorial peer review is the gateway to scientific publication. It was established to ensure that research papers were vetted by independent experts before they are published. Despite the importance of this process, its impact is still considered suboptimal and it needs to be improved. For this purpose, we need appropriate outcomes, particularly a validated tool that clearly defines the quality of peer review reports.

Objectives

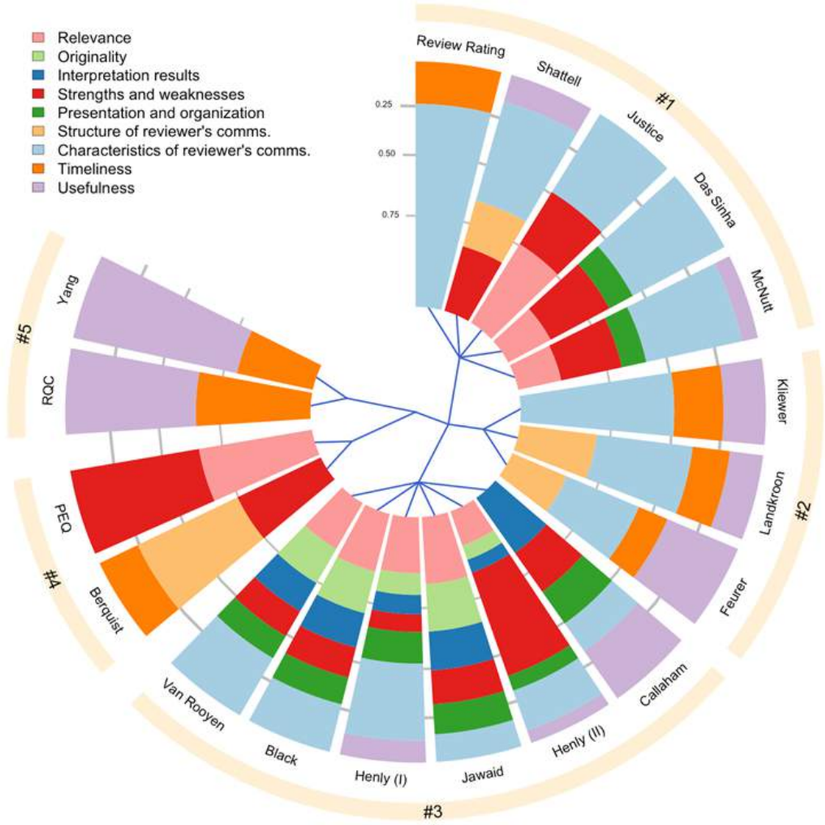
We aim to develop and validate a new tool to assess the quality of peer review reports in biomedical research

Methods

We performed a systematic review to identify and describe the existing tools used to assess peer review report quality. A steering committee composed by five members with different expertise defined the quality as "the extent to which a peer review report helps editors make a fair decision and authors improve the quality of the submitted manuscript". We conducted an online survey intended for biomedical editors and authors to 1) determine if participants endorse the proposed definition of peer review report quality; 2) identify the most important items to include in the tool; and 3) identify any missing items. Based on the participants' qualitative and quantitative answers, the steering committee reviewed all items and, ultimately, drafted and refined the final version of the tool. Lastly, we evaluate the psychometric properties of the tool.

Results

The systematic review allowed identifying 24 tools aimed at assessing the quality of peer review reports: none reported any definition of peer review report quality, only one described the scale development, and 10 provided measures of reliability and validity. Further, the development and validation process resulted from a small consensus of people, and the concepts evaluated by these tools were quite heterogeneous, as it is shown in the figure next page (1). A total of 446 biomedical editors and authors participated in the online survey. The majority of participants (84%) agreed on the definition of peer review report quality we proposed. The initial items to assess peer review report quality included in the survey questionnaire were generally highly rated with a mean score ranging from 3.38 (SD=1.13) to 4.60 (SD=0.69) (scale 1 to 5). Participants suggested 13 items that were not included in the initial list of items. We finally developed ARCADIA (Assessment of Review reports with a Checklist Available to eDItors and Authors). The tool is a checklist that includes five domains and 14 items (2). We are currently validating the tool.



Hierarchical clustering of tools based on the nine quality domains. The figure shows which quality domains are present in each tool. A slice of the chart represents a tool, and each slice is divided into sectors, indicating quality domains (in different colours).
The area of each sector corresponds to the proportion of each domain within the tool.

Conclusions

Several tools are available to assess the quality of peer review reports; however, the development and validation process is questionable and the concepts evaluated by these tools vary widely. ARCADIA is the first checklist that has been systematically developed to assess the quality of peer review reports. Its development is based on an exhaustive systematic review and the perspectives of a large sample of editors and authors. It could be used regularly by editors to evaluate the reviewers' work, and also as an outcome when evaluating interventions to improve the peer review process.

1. Superchi C, González JA, Solà I, Cobo E, Hren D, Boutron I. Tools used to assess the quality of peer review reports: a methodological systematic review. BMC Medical Research Methodology. 2019;19(1):48. doi: 10.1186/s12874-019-0688-x.
2. Superchi C, Hren D, Blanco D, Rius R, Recchioni A, Boutron I, González JA. The development of ARCADIA: a tool for assessing the quality of peer review reports in biomedical research. (submitted)

List of publications

Articles published/accepted

1. **Bertizzolo L**, Bossuyt PM, Atal I, Ravaud P, Dechartres A. Disagreements in risk of bias assessment for randomised controlled trials included in more than one Cochrane systematic reviews: a research on research study using cross-sectional design. BMJ Open. 2019;9(4):e028382.
2. **Biggane AM**, Olsen M, Williamson PR. PPI in research: a reflection from Early Stage Researchers. Research Involvement and Engagement [Accepted October 2019]
3. **Biggane AM**, Williamson PR, Ravaud P, Young B. Participating in core outcome set development via Delphi surveys: Qualitative interviews provide pointers to inform guidance. BMJ Open [Accepted October 2019]
4. **Biggane AM**, Brading L, Ravaud P, Young B, Williamson PR. Survey indicated that core outcome set development is increasingly including patients, being conducted internationally and using Delphi surveys. Trials. 2018;19(1):113.
5. **Blanco D**, Altman D, Moher D, Boutron I, Kirkham JJ, Cobo E. Scoping review on interventions to improve adherence to reporting guidelines in health research. BMJ Open. 2019 May 9;9(5):e026589.
6. **Blanco D**, Biggane AM, Cobo E, and the MiRoR network. Are CONSORT checklists submitted by authors adequately reflecting what information is actually reported in published papers? Trials. 2018 Jan 29;19(1):80.
7. **Blanco D**, Kirkham JJ, Altman DG, et al. Interventions to improve adherence to reporting guidelines in health research: a scoping review protocol. BMJ Open 2017;7:e017551.
8. **Ghannad M**, Yang B, Leeflang M, Aldcroft A, Bossuyt PM, Schroter S, Boutron I. Evaluating an editorial intervention to reduce spin in the abstract conclusion of manuscripts: a randomized controlled trial. Retrieved from osf.io/y2ewa (2019, August 6)
9. **Ghannad M**, Olsen M, Boutron I, Bossuyt PM. A systematic review finds that spin or interpretation bias is abundant in evaluations of ovarian cancer biomarkers. Journal of Clinical Epidemiology. 2019;116:9-17.
10. **Gkioni E**, Rius R, Dodd S, Gamble C. A systematic review describes models for recruitment prediction at the design stage of a clinical trial. Journal of Clinical Epidemiology. 2019;115:141-149.
11. **Glonti K,** Boutron I, Moher D, Hren D. Journal editors' perspectives on the roles and tasks of peer reviewers in biomedical journals: qualitative study. BMJ Open 2019;0:e033421. (Accepted)
12. **Glonti K**, Cauchi D, Cobo E, Boutron I, Moher D, Hren D. A scoping review on the roles and tasks of peer reviewers in the manuscript review process in biomedical journals. BMC Medicine. 2019;17(1):118.
13. **Glonti K**, Hren D. Editors' perspectives on the peer-review process in biomedical journals: protocol for a qualitative study. BMJ Open 2018;8:e020568.
14. **Glonti K**, Cauchi D, Cobo E, et al. A scoping review protocol on the roles and tasks of peer reviewers in the manuscript review process in biomedical journals. BMJ Open 2017;7:e017468.
15. **Koroleva A**, Kamath S, Paroubek P. Measuring semantic similarity of clinical trial outcomes using deep pre-trained language representations. Journal of Biomedical Informatics: X, 2019;4:100058.
16. **Nguyen VT**, Young B, Ravaud P, Naidoo N, Benchoufi M, Boutron I. Overcoming Barriers to Mobilising Collective Intelligence in Research: Qualitative Study of Researchers with Experience of Collective Intelligence, Journal of Medical Internet Research 2019;21(7):e13792.
17. **Nguyen VT**, Benchoufi M, Young B, Ghosn L, Ravaud P, Boutron I, A scoping review provided a framework for new ways of doing research through mobilising collective intelligence, Journal of Clinical Epidemiology, 2019;110:1-11.

18. **Norman CR**, Leeflang MMG, Porcher R, Névéol, A. Measuring the impact of screening automation on meta-analyses of diagnostic test accuracy. Systematic Reviews. 2019;8(1):243.
19. **Norman CR**, Gargon E, Leeflang MMG, Névéol A., Williamson PR. Evaluation of an automatic article selection method for timelier updates of the Comet core outcome set database. Database. Oxford University Press (Accepted)
20. **Nyanchoka L**, Tudur-Smith C, Thu VN, Iversen V, Tricco AC, Porcher R. A scoping review describes methods used to identify, prioritize and display gaps in health research. Journal of Clinical Epidemiology. 2019;109:99-110.
21. **Nyanchoka L**, Tudur-Smith C, Porcher R, Hren D. Key stakeholders' perspectives and experiences with defining, identifying and displaying gaps in health research: a qualitative study protocol. BMJ Open. 2019;9(8):e027926.
22. **Olarte Parra C**, Van de Bruaene C; Weynants L; Nagler EV; McAleenan A; Elbers RG; Higgins JP; Goetghebeur E. Pre-emptive versus non pre-emptive kidney transplantationfor end-stage kidney disease, Cochrane Library 2018
23. **Olsen M**, Ghannad M, Lok C, Bossuyt PM. Shortcomings in the evaluation of biomarkers in ovarian cancer: a systematic review. Clinical Chemistry and Laboratory Medicine. 2019. (In press)
24. **Olsen M**, Zhelev Z, Hunt H, Peters JL, Bossuyt PM, Hyde C. Use of test accuracy study design labels in NICE's diagnostic guidance. Diagnostic and Prognostic Research. 2019;3:17.
25. **Sharp MK**, Tokalić R, Gómez G, Wager E, Altman DG, Hren D. A cross-sectional bibliometric study showed suboptimal journal endorsement rates of STROBE and its extensions. Journal of Clinical Epidemiology 2019;107:42-50.
26. **Sharp MK**, Bertizzolo L, Rius R, Wager E, Gómez G, Hren D. Using the STROBE statement: survey findings emphasized the role of journals in enforcing reporting guidelines. Journal of Clinical Epidemiology. 2019;116:26–35.
27. **Sharp MK**, Utrobičić A, Gómez G, et al. The STROBE extensions: protocol for a qualitative assessment of content and a survey of endorsement. BMJ Open 2017;7:e019043.
28. **Superchi C**, González JA, Solà I, Cobo E, Hren D, Boutron I. Tools used to assess the quality of peer review reports: a methodological systematic review. BMC Medical Research Methodology. 2019;19(1):48.
29. **Vo TT**, Porcher R, Chaimani A, Vansteelandt, S. A novel approach for identifying and addressing case-mix heterogeneity in individual patient data meta-analysis. Research Synthesis Methods 2019 (In press)
30. **Vo TT**, Superchi C, Boutron I, Vansteelandt S. The conduct and reporting of mediation analysis in recently published randomized controlled trials: results from a methodological systematic review. Journal of Clinical Epidemiology. 2019;117:78-88.

Conference papers

1. **Koroleva A**, Paroubek P. Extracting relations between outcome and significance level in Randomized Controlled Trials (RCTs) publications. In: Proceedings of the 18th BioNLP Workshop and Shared Task, Florence (Italy), 2019 (pp. 359–369).
2. **Koroleva A**, Paroubek P. Annotating Spin in Biomedical Scientific Publications: the case of Randomized Controlled Trials (RCTs). In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), Miyazaki (Japan), 2018 (pp. 359–369)
3. **Koroleva A**. Vers la détection automatique des affirmations inappropriées dans les articles scientifiques. In: 18e REncontres jeunes Chercheurs en Informatique pour le TAL (RECITAL 2017), 2017 (pp. 135 – 148)
4. **Koroleva A**, Paroubek P. Automatic detection of inadequate claims in biomedical articles: first steps. In: Proceedings of Workshop on Curative Power of MEdical Data, Constanta (Romania), 2017

5. **Koroleva A**, Paroubek P. On the contribution of specific entity detection and comparative construction to automatic spin detection in biomedical scientific publications. In: Proceedings of The 2nd Workshop on Processing Emotions, Decisions and Opinions (EDO 2017)
6. **Norman CR**, Leeflang M, Zweigenbaum P, Névéol A. Tri Automatique de la Littérature pour les Revues Systématiques. In: 24e Conférence sur le Traitement Automatique des Langues Naturelles (TALN) 2017 (p. 234)
7. **Norman CR**. A distantly supervised dataset for automated data extraction from diagnostic studies. In: Proceedings of the 18th BioNLP Workshop and Shared Task, Florence (Italy), 2019 (pp. 105– 114).
8. **Norman CR**, Leeflang M, Zweigenbaum P, Névéol A. Automating Document Discovery in the Systematic Review Process: How to Use Chaff to Extract Wheat". In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC), 2018 (pp. 3681-3687)
9. **Norman CR**, Leeflang M, Névéol A.Data Extraction and Synthesis in Systematic Reviews of Diagnostic Test Accuracy: A Corpus for Automating and Evaluating the Process. Proceedings of the AMIA Annual Symposium, 2018 (pp. 817-826)

Articles submitted/under review

1. **Bertizzolo L**, Bossuyt P, Atal I, Ravaud P, Dechartres A. With an incomplete or unclear report, studies can be classified at high or low risk of bias: a research on research study. (Submitted)
2. **Blanco D**, Hren D, Kirkham JJ et al. A survey exploring biomedical editors' perceptions of editorial interventions to improve adherence to reporting guidelines. F1000Research 2019, 8:1682. https://doi.org/10.12688/f1000research.20556.1 [version 1; peer review: awaiting peer review].
3. **Gkioni E**, Dodd S, Rius R, Gamble C. Statistical models to predict recruitment in clinical trials are rarely used by statisticians in UK and European networks. (Under review)
4. **Glonti K**, Boutron I, Moher D, Hren D. Journal editors' perspectives on the communication practices in biomedical journals: a qualitative study. (Submitted)
5. **Superchi C**, Hren D, Blanco D, Rius R, Recchioni A, Boutron I, González JA. The development of ARCADIA: a tool for assessing the quality of peer review reports in biomedical research. (Submitted)
6. **Vo TT**, Porcher R, Vansteelandt S. Assessing the impact of case-mix heterogeneity in meta-analysis: novel use of i2 statistic and prediction interval. (Submitted)

Commentaries and letters to editors

1. **Gkioni E**, Glonti K, Dodd S, et al. DIABRISK-SL trial: further consideration of age and impact of imputations. BMC Medicine 17, 121 (2019)
2. **Koroleva A**, Olarte Parra C, Paroubek P. On improving the implementation of automatic updating of systematic reviews. JAMIA Open 2019
3. **Nguyen VT**, Superchi C, Boutron I, 2-Year outcome from two parallel randomized controlled trials. Reporting considerations. Osteoarthritis and Cartilage 27 2019 e3-e4
4. **Norman C**, Nguyen TV, Névéol A. Contribution of Natural Language Processing in Predicting Rehospitalization Risk. Med Care. 2017;55(8):781.
5. **Olsen M**, Sharp MK, Bossuyt PM. From the theoretical to the practical: how to evaluate the ethical and scientific justifications of randomized clinical trials. Journal of Clinical Epidemiology. 2018;99:170-171.
6. **Olarte Parra C**, Nyanchoka L, Goethghebeur E. Different Pediatric Survival After Cardiac Arrest. Jama Pediatrics 2017;171(7):708-709.
7. **Sharp MK**, Hren D, Altman D. The STROBE Extensions: Considerations for Development. Epidemiology. 2018 29(6):e53–e56.
8. **Sharp MK**, Nyanchoka L, Hren D. Every ROSE has its thorns. Environmental Evidence 7, 20 (2018)

The MiRoR project leading to this publication has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 676207

www.miror-ejd.eu
contact@miror-ejd.eu