University of Basel
Department of Clinical Research

University Hospital Basel

IMPROVING VALUE OF CLINICAL RESEARCH –

AN EVIDENCE-BASED APPROACH

Inaugural dissertation

to be awarded the degree of

Dr. sc. med.

presented at the

Faculty of Medicine

of the University of Basel

by

Belinda von Niederhäusern

from Grenchen, Solothurn

Basel, 2017

Approved by the Faculty of Medicine

On application of

Prof. Dr. med. Christiane Pauli-Magnus (first supervisor)

PD Dr. med. Matthias Briel (first supervisor)

Prof. Matthias Schwenkglenks (second supervisor)

Prof. Dr. med. Mirjam Christ-Crain (second supervisor)

Prof. Christian Burri (external expert)

Basel, September 18, 2017

Prof. Dr. Thomas C. Gasser, Dean

TABLE OF CONTENTS

ACKNOWLEDGMENTS

Specifically, I would like to take the opportunity to thank Prof. Christiane Pauli-Magnus and PD Dr. Matthias Briel for their tremendous support at all stages of this thesis. It was a great pleasure to be supported by visionaries who are passionate and convinced in what they do. Both were highly supportive and allowed me to investigate, develop, and grow, both scientifically and personally, in an environment that endorsed trial, creativity, and error. At the same time, they shed light on paths that I would have never imagined to walk on during my PhD. I will forever be grateful for this opportunity and mostly, their trust in me as their first PhD student at the Clinical Trial Unit. Prof. Pauli-Magnus and PD. Dr. Briel have not only been supervisors, but became role models, and will be mentors for the future.

I would also like to thank Prof. Matthias Schwenkglenks for his great support with this thesis, specifically all aspects related to costs. I very much enjoyed working for and with him due to his generous advice, tremendous knowledge, and his pleasant personality. I also thank Prof. Mirjam Christ-Crain who has always supported me in all aspects related to this PhD. Further, I would like to thank Prof. Christian Burri for accepting to be the external expert of this thesis. Whenever we met in person, he provoked new thoughts, was positive and supportive of the work I do, and finally, made me laugh.

I want to thank all of you who devoted time, passion, and effort to any of the projects mentioned in here, including all co-authors. In particular, Prof. Gordon Guyatt for his stimulating advice, Dr. Benjamin Speich, Dr. Nadine Schur, and Dr. Priya Satalkar for fruitful discussions, fantastic team work, and their friendship. The colleagues at the Clinical Trial Unit deserve a special mention for their endless collaborative support, notably the statistics, data management, monitoring, and application development groups, but also all study nurses for a very cheerful working environment.

A special gratitude goes out to the 52 experts who devoted their hours and thoughts to the development of our quality work until consensus was reached. Without them, our framework would have been a meaningless theory. Further, I would like to thank the Swiss Clinical Trial Organization for their support with everything related to quality and patient engagement, and the Swiss Office of Public Health that funded some of my research and was supportive in all aspects.

To my friends outside of University and those who crossed my path during the PhD, I am forever grateful for your friendship and support through good and rough times.

Finally, I specifically thank my parents, Elisabeth and Hans-Rudolf von Niederhäusern, and my partner, Florian Stroh, who have always believed in my endeavors and never stopped supporting my life choices. You all provided me with the backbone I needed to be successful, celebrated each little step with joy and laughter, always stood by my side when I was in doubt, and made me the person I am today.

PLAIN SUMMARY

Evidence suggests that 85% of biomedical research spending, i.e. 200 billion US dollars every year, goes to waste. In 2014, *The Lancet* published a series of five reviews showing how dividends from the investment in research might be increased at five stages - from the relevance and priorities of the questions being asked, to how the research is designed, conducted, and reported. Value and waste have since then become buzzwords in the academic as well as public debate surrounding health research. Although academic institutions are the major driving force of patient-oriented clinical research receiving large proportions of public funding they have been slow responders to the Series' recommendations. Some of the identified underlying reasons include a lack of a common understanding of "value" as a concept and sparse practical guidance for academia on how to improve it. This work represents the first effort to formulate an academic response to *The Lancet* series on increasing value in clinical research by investigating the two distinct concepts in the equation: "Quality" and "cost".

In a first step, we systematically reviewed existing quality concepts, both in the medical literature and across international clinical research stakeholder groups. Precise definitions of quality were sparse, and stakeholder perspectives of crucial components of quality varied. Based on these findings, we then engaged international stakeholder representatives in the creation of a comprehensive, consensus-based framework for the quality of clinical research that is applicable to all study types and spans the entire lifecycle of a clinical study, i.e. from conceptualization of the research question to dissemination of study results. Primarily, it is designed to be operationalized in the academic setting and fully supports the REWARD Statement. This framework builds the foundation for a common understanding of the concept of "quality" and its practical assessment. At Swiss national level, the framework has triggered all stakeholders to convene in a first symposium on how to increase value of academic clinical research and serves as an agenda for future research on research.

In a second step, we systematically reviewed the current evidence on the costs and associated resource use of Randomized Controlled Trials (RCTs), which we found to be sparse. Based on this, we laid the foundation for future study cost assessments in academia by (i) developing a comprehensive list of items for the retrospective and prospective assessment of costs, and (ii) generating first empirical evidence on main cost drivers in a case report on two academic RCTs. Although these two RCTs were conducted in very different settings and resulted in vastly different costs, the main drivers, i.e. personnel costs

during conduct phase, were the same. In addition, we investigated the added value of two innovative aspects that affect both study quality and cost, i.e. risk-based trial monitoring and remote data collection. We show that both concepts may increase the cost-effectiveness of trial conduct and thereby increase value, but only if the methodology is further investigated and then, rigorously implemented.

Although we did not take the initially envisioned cost-consequence approach, we have certainly created awareness on value and waste in the academic context and engaged the major stakeholders in fundamental discussions on how to improve the current situation. In the future, the costs occurred need to inform quality assessments of clinical studies in order to create a tool that creates “value”, rather than sole quality conformity. Furthermore, the willingness-to-pay of academic decision makers in resource-constrained settings will weigh into the value equation and needs further investigation in the future. The impact of this work - and whether it eventually increases value in the system - now critically depends on its rigorous implementation, evaluation, and refinement.

CHAPTER 1

INTRODUCTION

"Perhaps all of us engaged in the enterprise we call "science" need to pause and reflect on the present state of what we do."

(Kleinert & Horton, The Lancet, 2014)

1.1 The need for improving value and reducing waste in clinical research

The biomedical research complex has been estimated to consume almost a quarter of a trillion US dollars every year. Unfortunately, evidence suggests that a high proportion of this sum is avoidably wasted [1-7]. The output of new pharmaceutical drugs has been decreasing for the past decade and the prices have risen steadily, leading to access problems for many patients [8]. Most university-initiated technology transfer units created to protect and sell academically generated intellectual property cost more than they earn [9]. Although much of the world follows the Moore law on the doubling of output (e.g. computing power) per unit cost every 2 years, drug development and clinical trials are moving in the opposite direction [10].

In 2014, The Lancet published a series of five reviews showing how dividends from the investment in research might be increased at five stages - from the relevance and priorities of the questions being asked, to how the research is designed, conducted, and reported [1, 2, 4-7]. Seventeen recommendations were addressed to five main stakeholders – funders, regulators, journals, academic institutions, and researchers. In the same year, Moors et al. [8] suggested a combination of reforms including technological and organizational changes, changes in the regulatory, patent, and reimbursement system, and social and/or political changes to make drug development more sustainable.

In 2016, a follow-up review provided some initial observations of the possible effects of the Lancet Series [11]. It suggested that some movement had been provoked across stakeholders, but that still much more needed to be done to effectively increase value and reduce waste across the biomedical research system. One of the leaders of the movement, John Ioannidis, said that *"(...) not only are most research findings false, but, furthermore, most of the true findings are not useful"* [12]. Interestingly, academic institutions had been among the slowest responders to the Series' recommendations. As the major driving force of patient-oriented clinical research receiving large proportions of public funding [13, 14] and producing the majority of scientific publications in this area [15], academic institutions would be ideally placed to lead the movement. Historically, however, academia had been criticized for low quality research [16]. Criticism included allegations of financial conflicts of interest [17], scientific misconduct by a few investigators[18], low dissemination rates of clinical trial results [19], and a significantly higher risk for discontinuation compared to industry-funded trials [20]. Reasons for the slow progress may be that academia is a complex ecosystem including many stakeholders with different agendas and a lack of common policies [11].

In summary, the evidence that value of research needs to be improved is compelling – particularly for academia. Still, the definition of "value" remains vague. This is particularly challenging because "value" in itself is the result of an equation of "worth", i.e. "quality" versus "monetary investment" or "cost", and "willingness-to-pay". In order to be able to successfully operationalize the Lancet Series' recommendations and thereby increase "value", however, one needs to be very specific about these two underlying concepts.

In this PhD work, we first aimed to develop a cost-consequence approach to improving the quality of research. However, we early-on identified gaps in the evidence-base relating to the concepts underlying this approach, i.e. (1) a common understanding of the concept of "quality" of clinical research, its definition, and its practical assessment, and (2) an empirical evidence base for the "cost" of clinical research, its components, and main drivers. In the following paragraphs, I describe the rationale and the approach we took to fill both of these gaps. Finally, I highlight the rationale for two projects that we conducted to assess the cost-effectiveness of two aspects of high quality research, risk-based monitoring and remote data collection.

1.2 The need for a common definition of the quality of research

"There is clearly a strong feeling among many scientists, and not only Nobel Prize Winners, that something has gone wrong with our system for assessing the quality of scientific research."

(Kleinert & Horton, The Lancet, 2014)

Approaches to assess or measure the quality of single clinical studies or whole research programs have been limited, mono-dimensional, and often criticized for not being sufficient to tackle waste. Quality assurance measures, such as full compliance with the International Conference on Harmonization of Good Clinical Practice (ICH GCP) requirements or 100% Source Data Verification were even deemed major sources of waste leading to high costs in the system without proven benefit [21-23]. The existing international quality guidelines, GCP, were criticized for their non-scientific development process and a lack of consensus across stakeholder groups [24, 25]. In addition, they were deemed an unsuitable standard for investigator-initiated clinical research [26, 27].

In economics theory, disputes on a definition for "product quality" have a long-standing tradition: *"Quality is a complex and multifaceted concept. It is also the source of great confusion: managers –particularly those in in different functions- frequently fail to communicate precisely what they mean by the term"* [28]. Quality theories range from "user-based" to "product-based" to "manufacturing-based". Equivalents in clinical research could be "patient-centered" to "methods-based" to "operational" quality of a study. The "value-based" approach defines quality in terms of costs and prices. According to this view, a quality product is one that provides performance at an acceptable price or conformance at an acceptable cost, which has become more prevalent as "quality is increasingly apt to be discussed and perceived in relationship to price" [28].

In clinical research, perspectives, priorities, and incentives concerning research quality naturally vary across the different involved stakeholders. Manuscript I entitled "*Towards the development of a comprehensive framework: Qualitative systematic survey of definitions of clinical research quality*" describes our systematic search for the existing quality concepts, definitions and criteria across clinical research stakeholders, both in the published literature and on stakeholder group websites. This work aimed to reflect the heterogeneous quality landscape, identify divergent and common stakeholder perceptions, and lay the foundation for the future development of a comprehensive definition of quality.

In his publication on "how to make more published research true" in 2014, John Ioannidis suggested that "*joint efforts by multiple stakeholders (in biomedical research) may yield solutions that are more likely to be more widely adopted and thus successful"* [29]. We fully embraced such a user-centered approach by engaging over 100 international experts from seven stakeholder groups in a four-round collaborative consensus finding process to define the critical items for the quality of clinical research. Manuscript II entitled "*Towards increasing value and reducing waste in academic clinical research: Consensus on a comprehensive framework of clinical research quality"* describes the consensus-based development of our framework and highlights first applications in the Swiss context.

1.3 The need for evidence on the cost of research

Cost estimates for research and development (R&D) for new drugs in 2009 ranged from USD 92 million to USD 884 million per compound [30, 31]. The primary driver of the rising costs is clinical costs, especially clinical trials, which increased 10-fold from 1991 to 2003 [32, 33]. With a number of initiatives and regulations that were implemented to improve research quality and to increase participant protection [34], the complexity and the administrative burden of RCTs increased, again raising their overall costs [35-37]. Ultimately, the number of RCTs has decreased over the last decade [38] and a substantial proportion of initiated RCTs are prematurely discontinued due to organizational and recruitment problems [39] risking that more uncertainties about medical treatments will go unaddressed.

While efforts to make clinical trials more cost-efficient are urgently needed, several institutions have criticized the published total cost estimates to be "intransparent" or "potential exaggerations" to justify high drug prices [40-42]. A pre-requisite for an optimization process, however, are reliable empirical cost data and evidence on cost-drivers. Therefore, we aimed to generate an evidence base by a) systematically compiling the existing evidence on cost and resource use in RCTs, and b) retrospectively collecting detailed resource use and associated costs of two RCTs conducted within our close network. Manuscript III entitled "*Systematic review on costs and resource use of randomised clinical trials shows a lack of empirical data"* provides the results of our systematic search for empirical evidence on clinical trial costs. Manuscript IV entitled "*Cost and resource use evaluation of randomised clinical trials: a case study exemplifying standardised assessment using a comprehensive cost item list"* describes the first case report of detailed cost and resource use estimates of two RCTs.

1.4 The need for cost-effective solutions

1.4.1 Approach 1: Trial monitoring

The cost of assuring operational "quality" in clinical trials – such as monitoring the compliance with complex regulatory guidelines – has been described to be one of the major drivers of exploding R&D expenditure – and waste in the system [43, 44]. Financial estimates of a single monitoring site visit range from USD 800 in 1991 to USD 1500 in 2009 [45, 46], with conservative cost estimates for one single query of USD 150 [47]. Traditional monitoring approaches relied on intensive on-site visits and 100% Source Data Verification (SDV) irrespective of the risk levels in the study, leading to high cost and only limited contribution to clinical trial data quality [21-23]. Recent developments at international bodies and regulatory agencies such as ICH and the European Medicines Agency have supported the need for risk-proportionate approaches to clinical trial monitoring and published respective guidance [48-52]. In the academic setting, restricted resources often oblige investigators to apply a risk-based approach to trial monitoring. The Risk ADApted MONitoring (ADAMON) Project proposed a first instrument for the risk analysis of on-site monitoring in the academic setting [53], which we follow at the Clinical Trial Unit Basel since 2012.

In order to shed light on "what works and what doesn't", we conducted the first comprehensive retrospective study assessing the cost and potential benefits of our current monitoring approach. In manuscript V entitled "*Generating evidence on a risk-based monitoring approach in the academic setting - Lessons learned"* we provide evidence on the characteristics of findings documented during on-site visits, the factors that might influence the number and types of monitoring findings, the costs associated with our approach, and the experience of our monitors with the risk-based approach.

1.4.2 Approach 2: High quality data

High quality research relies on the collection of high quality data. Traditionally, this is done in the inpatient setting or through ambulatory visits to a study site. The widespread availability of new technologies has the potential of shifting some research activities, including enrollment, managing trial activity, reporting results, and safety oversight, away from study sites. Such "remote" research may encourage the participation of a more diverse group of patients in research with improved recruitment rates and at lower costs than those of conventional trials [1-3], and puts individuals, rather than investigative sites, at the center of the research process. However, issues around retention and data quality remain [10, 54-59]. Therefore, a combination with direct interactions with the research team may allow remote approaches to be optimally leveraged [8, 9]. While increasing interest and support from

regulators, sponsors, and patients has created much “buzz” around these trials, the methodology is still in early stages of development and requires further investigation. Manuscript VI entitled “*Validity of mobile electronic data capture in clinical studies: A pilot study in a pediatric population”* describes our pilot study (the TOMACHI study) investigating the feasibility of remotely collecting valid (i.e. complete and correct) clinical data and samples in a pediatric population utilizing mobile technologies. In addition, we assessed the general acceptance, reasons for non-consent, and the resulting costs of this study. As a model, we chose children and their caregivers as a population that is a) in urgent need for innovative clinical study designs to advance the current knowledge on dosing and action of routinely used medicines [60-62] and b) familiar with electronic technologies.

1.5 Main Objectives of this PhD

1) Develop a common understanding of the quality of clinical research across stakeholders
2) Create an evidence base for the costs occurred in clinical research, and their main drivers
3) Evaluate the cost-effectiveness of two aspects of high quality research, risk-based monitoring and remote data collection

1.6 Contributions by the PhD student

I had the great opportunity to be the first PhD student at the Clinical Trial Unit (CTU) and was therefore part of the conceptualization of this work from the very onset. In this fostering environment, I was encouraged to come up with own ideas and received the freedom to develop new study plans in very close collaboration with my supervisors. I therefore had a substantial role in all aspects of the studies mentioned in this dissertation, i.e. from design, planning and conduct trough analysis and dissemination of study results.

After conceptualization and identification of the most suitable study designs with my supervisors, I drafted the first version of the study protocols. This included, depending on the study, the design of systematic literature reviews and reviews of websites, endpoint definitions, quantitative or qualitative analyses plans for main and secondary objectives, data management plans, monitoring plans, and the requirements plan for a mobile application (TOMACHI). I coordinated the different teams involved in each study including statisticians, data managers, monitors, study nurses, and app developers during the drafting process, and

submitted ethics proposals (and amendments) for the TOMACHI study, and ethics waivers for the quality framework study and the monitoring study.

In the conduct phase, I was responsible for data collection, the project management and coordination of the teams contributing to these projects, ranging from 3-12 members. In the two systematic reviews (of which one was published in PLOS One, and one is currently under review at The Journal of Clinical Epidemiology) I coordinated and extracted data with the great help of many team members. Then, after iterative discussion with my supervisors, I set up a first matrix for the quality framework which I circulated across over 100 international experts in a Delphi process. I planned and coordinated the Delphi process, collected the data and engaged with all Delphi participants over two years. In the cost projects, I contributed to the development of a comprehensive cost item list and the cost interviews with principal investigators. In the TOMACHI study, I was the designated study coordinator overlooking all activities that were conducted on site (i.e. recruitment, data collection) by a study nurse team and a recruiting physician, and managed all interactions with the ethics committee. I was responsible for data management of all qualitative and quantitative data, except the TOMACHI study which required data transfer from a mobile application to the data base SecuTrial. Throughout these studies, I established a network across stakeholders (e.g. The Federal Office of Public Health, EUPATI, or the Swiss Clinical Trial Organization) who actively engaged in and supported my work. Additionally, I helped drafting other study protocols (e.g. SPIRIT, MARTA) in our collaborative group, collected data for other projects, and gave critical and constructive feedback on manuscript drafts.

In the analysis phase, I analyzed both qualitative and descriptive quantitative data, except for the TOMACHI study which was performed by a CTU statistician. I critically interpreted the data together with my supervisors and co-authors and developed first drafts for all manuscripts, coordinated the critical revision by co-authors, submitted and revised manuscripts as first and co-first author, and presented and discussed our work at international and national conferences.

Finally, although my PhD position was fully funded by the CTU, I wrote a funding proposal to the PhD Program for Health Sciences (PPHS) for a top-up stipend which was granted. In addition to my PhD work, I also had the fantastic opportunity to take over different roles at the CTU, e.g. as maternity cover. For example, I coordinated parts of the Diploma of Advance Studies course and supervised 12 students during their Diploma theses, and managed the roll-out of a pilot study investigating the feasibility of electronic general consent on clinical wards. I am deeply grateful for these diverse opportunities in a very supportive

environment that supported my learning, challenged my skills, and taught me to be an independent researcher working in and with an interdisciplinary environment.

References

[1] Al-Shahi Salman R, Beller E, Kagan J, Hemminki E, Phillips RS, Savulescu J, et al. Increasing value and reducing waste in biomedical research regulation and management. Lancet. 2014;383:176-85.
[2] Chalmers I, Bracken MB, Djulbegovic B, Garattini S, Grant J, Gulmezoglu AM, et al. How to increase value and reduce waste when research priorities are set. Lancet. 2014;383:156-65.
[3] Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. Lancet. 2009;374:86-9.
[4] Chan AW, Song F, Vickers A, Jefferson T, Dickersin K, Gotzsche PC, et al. Increasing value and reducing waste: addressing inaccessible research. Lancet. 2014;383:257-66.
[5] Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, Julious S, et al. Reducing waste from incomplete or unusable reports of biomedical research. Lancet. 2014;383:267-76.
[6] Ioannidis JP, Greenland S, Hlatky MA, Khoury MJ, Macleod MR, Moher D, et al. Increasing value and reducing waste in research design, conduct, and analysis. Lancet. 2014;383:166-75.
[7] Macleod MR, Michie S, Roberts I, Dirnagl U, Chalmers I, Ioannidis JP, et al. Biomedical research: increasing value, reducing waste. Lancet. 2014;383:101-4.
[8] Moors EH, Cohen AF, Schellekens H. Towards a sustainable system of drug development. Drug Discov Today. 2014;19:1711-20.
[9] Lalkaka R. Technology business incubators to help build an innovation-based economy. Journal of Change Management. 2002;3:167-76.
[10] Dorsey ER, Venuto C, Venkataraman V, Harris DA, Kieburtz K. Novel Methods and Technologies for 21st-Century Clinical Trials: A Review. JAMA Neurol. 2015;72:582-8.
[11] Moher D, Glasziou P, Chalmers I, Nasser M, Bossuyt PM, Korevaar DA, et al. Increasing value and reducing waste in biomedical research: who's listening? Lancet. 2016;387:1573-86.
[12] Ioannidis JP. Why Most Clinical Research Is Not Useful. PLoS Med. 2016;13:e1002049.
[13] Dorsey ER, de Roulet J, Thompson JP, Reminick JI, Thai A, White-Stellato Z, et al. Funding of US biomedical research, 2003-2008. Jama. 2010;303:137-43.
[14] Moses H, 3rd, Matheson DH, Cairns-Smith S, George BP, Palisch C, Dorsey ER. The anatomy of medical research: US and international comparisons. Jama. 2015;313:174-89.
[15] Ioannidis JP, Boyack KW, Klavans R. Estimates of the continuously publishing core in the scientific workforce. PLoS ONE. 2014;9:e101698.
[16] Meador KJ. Decline of clinical research in academic medical centers. Neurology. 2015;85:1171-6.
[17] Lo B. Serving two masters--conflicts of interest in academic medicine. N Engl J Med. 2010;362:669-71.
[18] Binder R, Friedli A, Fuentes-Afflick E. The New Academic Environment and Faculty Misconduct. Acad Med. 2016;91:175-9.
[19] Chen R, Desai NR, Ross JS, Zhang W, Chau KH, Wayda B, et al. Publication and reporting of clinical trial results: cross sectional analysis across academic medical centers. Bmj. 2016;352.
[20] Kasenda B, von Elm E, You J, Blumle A, Tomonaga Y, Saccilotto R, et al. Prevalence, characteristics, and publication of discontinued randomized trials. Jama. 2014;311.
[21] Tudur Smith C, Stocken DD, Dunn J, Cox T, Ghaneh P, Cunningham D, et al. The value of source data verification in a cancer clinical trial. PLoS ONE. 2012;7:e51623.
[22] Tudur Smith C, Williamson P, Jones A, Smyth A, Hewer SL, Gamble C. Risk-proportionate clinical trial monitoring: an example approach from a non-commercial trials unit. Trials. 2014;15:127.
[23] Baigent C, Harrell FE, Buyse M, Emberson JR, Altman DG. Ensuring trial validity by data quality assurance and diversification of monitoring methods. Clin Trials. 2008;5:49-55.
[24] Woolf SH. Practice guidelines, a new reality in medicine. II. Methods of developing guidelines. Arch Intern Med. 1992;152:946-52.

[25] Grimes DA, Hubacher D, Nanda K, Schulz KF, Moher D, Altman DG. The Good Clinical Practice guideline: a bronze standard for clinical research. Lancet. 2005;366:172-4.
[26] Stewart PM, Stears A, Tomlinson JW, Brown MJ. Regulation--the real threat to clinical research. Bmj. 2008;337:a1732.
[27] McMahon AD, Conway DI, Macdonald TM, McInnes GT. The unintended consequences of clinical trials regulations. PLoS Med. 2009;3:e1000131.
[28] Garvin DA. What Does "Product Quality" Really Mean? MIT Sloan Management Review. 1984;26.
[29] Ioannidis JP. How to make more published research true. PLoS Med. 2014;11:e1001747.
[30] Morgan S, Grootendorst P, Lexchin J, Cunningham C, Greyson D. The cost of drug development: a systematic review. Health Policy. 2011;100:4-17.
[31] Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, Lindborg SR, et al. How to improve R&D productivity: the pharmaceutical industry's grand challenge. Nat Rev Drug Discov. 2010;9:203-14.
[32] DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: New estimates of R&D costs. J Health Econ. 2016;47:20-33.
[33] Canada Clinical Trial Summit. Canadian Institute for Health Research. 2011.
[34] Pei H, Wu Y, Wei Y, Yang Y, Teng S, Zhang H. Remote ischemic preconditioning reduces perioperative cardiac and renal events in patients undergoing elective coronary intervention: a meta-analysis of 11 randomized trials. PLoS ONE. 2014;9:e115500.
[35] Sinha G. European move affects academic trials research. J Natl Cancer Inst. 2006;98:1100-1.
[36] Duley L, Antman K, Arena J, Avezum A, Blumenthal M, Bosch J, et al. Specific barriers to the conduct of randomized trials. Clin Trials. 2008;5:40-8.
[37] Eisenstein EL, Collins R, Cracknell BS, Podesta O, Reid ED, Sandercock P, et al. Sensible approaches for reducing clinical trial costs. Clin Trials. 2008;5:75-84.
[38] Bosch X. Europe's restrictive rules strangling clinical research. Nat Med. 2005;11:1260-.
[39] Kasenda B, von Elm E, You J, Blumle A, Tomonaga Y, Saccilotto R, et al. Prevalence, characteristics, and publication of discontinued randomized trials. JAMA. 2014;311:1045-51.
[40] Light DW, Warburton R. Demythologizing the high costs of pharmaceutical research. The London School of Economics and Political Science. 2011; 1745-8552 BioSocieties 1–17. www.palgrave-journals.com/biosoc/ (accessed August 25, 2016).
[41] Adams CP, Brantner VV. Estimating the cost of new drug development: is it really 802 million dollars? Health Aff (Millwood). 2006;25:420-8.
[42] Zimmer AS, Steeg PS. Meaningful prevention of breast cancer metastasis: candidate therapeutics, preclinical validation, and clinical trial concerns. J Mol Med. 2015;93:13-29.
[43] Olsen R, Bihlet AR, Kalakou F, Andersen JR. The impact of clinical trial monitoring approaches on data integrity and cost-a review of current literature. Eur J Clin Pharmacol. 2016;72:399-412.
[44] Bakobaki J, Joffe N, Burdett S, Tierney J, Meredith S, Stenning S. A systematic search for reports of site monitoring technique comparisons in clinical trials. Clin. 2012;9:777-80.
[45] Vantongelen K, Steward W, Blackledge G, Verweij J, Van Oosterom A. EORTC joint ventures in quality control: treatment-related variables and data acquisition in chemotherapy trials. Eur J Cancer. 1991;27:201-7.
[46] Rosa C, Campbell A, Kleppinger C, Sampson R, Tyson C, Mamay-Gentilin S. Quality assurance of research protocols conducted in the community: the National Institute on Drug Abuse Clinical Trials Network experience. Clin Trials. 2009;6:151-61.
[47] Pronker E, Geerts BF, Cohen A, Pieterse H. Improving the quality of drug research or simply increasing its cost? An evidence-based study of the cost for data monitoring in clinical trials. Br J Clin Pharmacol. 2011;71:467-70.
[48] Bakobaki JM, Rauchenberger M, Joffe N, McCormack S, Stenning S, Meredith S. The potential for central monitoring techniques to replace on-site monitoring: findings from an international multi-centre clinical trial. Clin Trials. 2012;9:257-64.

[49] Journot V, Pignon JP, Gaultier C, Daurat V, Bouxin-Metro A, Giraudeau B, et al. Validation of a risk-assessment scale and a risk-adapted monitoring plan for academic clinical research studies--the Pre-Optimon study. Contemp Clin Trials. 2011;32:16-24.
[50] Risk based approach to monitoring: Guidance for Industry. US Food And Drug Administration FDA; 2013.
[51] International conference on harmonisation of technical requirements for registration of pharmaceuticals for human use, Intergrated Addendum to ICH E6(R1): Guideline for Good Clinical Practice E6R(2)
[52] European Commission, Risk Proportionate Approaches in Clinical Trials-Recommendations of the Expert Group on Clinical Trials for the Implementation of Regulation (EU) no 536/2014 on Clinical Trials on Medicinal Products for Human Use, 2014.
[53] Brosteanu O, Houben P, Ihrig K, Ohmann C, Paulus U, Pfistner B, et al. Risk analysis and risk adapted on-site monitoring in noncommercial clinical trials. Clin Trials. 2009;6:585-96.
[54] McAlindon T, Formica M, Kabbara K, LaValley M, Lehmer M. Conducting clinical trials over the internet: feasibility study. Bmj. 2003;327:484-7.
[55] Campbell KR. An apple a day: changing medicine through technology and engagement. Future Cardiol. 2015;11:259-60.
[56] Anguera JA, Jordan JT, Castaneda D, Gazzaley A, Arean PA. Conducting a fully mobile and randomised clinical trial for depression: access, engagement and expense. BMJ Innov. 2016;2:14-21.
[57] Caldwell PHY, Hamilton S, Tan A, Craig JC. Strategies for Increasing Recruitment to Randomised Controlled Trials: Systematic Review. PLOS Medicine. 2010;7:e1000368.
[58] Watson JM, Torgerson DJ. Increasing recruitment to randomised trials: a review of randomised controlled trials. BMC Med Res Methodol. 2006;6:34.
[59] Murray E, Khadjesari Z, White IR, Kalaitzaki E, Godfrey C, McCambridge J, et al. Methodological challenges in online trials. J Med Internet Res. 2009;11:e9.
[60] Shirkey HC. Therapeutic orphans--everybody's business. Ann Pharmacother. 2006;40:1174.
[61] Turner S, Nunn AJ, Fielding K, Choonara I. Adverse drug reactions to unlicensed and off-label drugs on paediatric wards: a prospective study. Acta Paediatr. 1999;88:965-8.
[62] Avenel S, Bomkratz A, Dassieu G, Janaud JC, Danan C. [The incidence of prescriptions without marketing product license in a neonatal intensive care unit]. Arch Pediatr. 2000;7:143-7.

CHAPTER 1

FIRST AUTHOR PUBLICATIONS

Manuscript I: Towards the development of a comprehensive framework: Qualitative systematic survey of definitions of clinical research quality

Belinda von Niederhäusern[1*], Stefan Schandelmaier[2,3], Marie Mi Bonde[1], Nicole Brunner[1], Lars G. Hemkens[2], Marielle Rutquist[1], Neera Bhatnagar[3], Gordon H. Guyatt[3], Christiane Pauli-Magnus[1¶], and Matthias Briel[2,3¶]

[1]Clinical Trial Unit, Department of Clinical Research, University Hospital Basel, Basel, BS, Switzerland

[2]Basel Institute for Clinical Epidemiology and Biostatistics, Department of Clinical Research, University Hospital Basel, Basel, BS, Switzerland

[3]Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, ON, Canada

[¶] CPM and MB are Joint Senior Authors

Status: Published

PLoS One. 2017 Jul 17;12(7):e0180635.

Abstract

Objective: To systematically survey existing definitions, concepts, and criteria of clinical research quality, both developed by stakeholder groups as well as in the medical literature. This study serves as a first step in the development of a comprehensive framework for the quality of clinical research.

Study Design and Setting: We systematically and in duplicate searched definitions, concepts and criteria of clinical research quality on websites of stakeholders in clinical research until no further insights emerged and in MEDLINE up to February 2015. Stakeholders included governmental bodies, regulatory agencies, the pharmaceutical industry, academic and commercial contract research organizations, initiatives, research ethics committees, patient organizations and funding agencies from 13 countries. Data synthesis involved descriptive and qualitative analyses following the Framework Method on definitions, concepts, and criteria of clinical research quality. Descriptive codes were applied and grouped into clusters to identify common and stakeholder-specific quality themes.

Results: Stakeholder concepts on how to assure quality throughout study conduct or articles on quality assessment tools were common, generally with no *a priori* definition of the term quality itself. We identified a total of 20 explicit definitions of clinical research quality including varying quality dimensions and focusing on different stages in the clinical research process. Encountered quality dimensions include ethical conduct, patient safety/rights/priorities, internal validity, precision of results, generalizability or external validity, scientific and societal relevance, transparency and accessibility of information, research infrastructure and sustainability. None of the definitions appeared to be comprehensive either in terms of quality dimensions, research stages, or stakeholder perspectives.

Conclusion: Clinical research quality is often discussed but rarely defined. A framework defining clinical research quality across stakeholders' individual perspectives is desirable to facilitate discussion, assessment, and improvement of quality at all stages of clinical research.

Introduction

Clinical research is necessary to advance our knowledge and practice of diagnosing and preventing diseases and treating patients. However, its complexity and the regulatory requirements have significantly increased over the last few years, requiring an ever-rising level of scientific, methodological, regulatory and organizational know-how [1]. Global clinical research involves billions of dollars and millions of people, yet it is often poorly planned, inefficient, or "not useful", leading to considerable waste of private and public funding [1-8]. Low quality research may not only result in misleading findings [9], but may also compromise safety and rights of patients.

The regulatory international "ethical and scientific quality standard for designing, conducting, recording and reporting trials" – the Good Clinical Practice (GCP) guideline developed by the International Conference on Harmonisation (ICH) aims to ensure that safety and rights of participants are protected and that trial data are credible [10, 11]. The GCP guideline is a widely disseminated and applied standard for the broad concept of clinical research quality. However, its limitations include development as an agreement between industry and regulatory experts and its focus on data accuracy and extensive formal requirements has been criticized as an unsuitable standard for investigator-initiated clinical research [12, 13]. The GCP guidelines lack a broad stakeholder consensus and a sound evidence-base [14, 15].

In academic clinical research, "quality" often relates to design and implementation from the standpoint of scientific rigor. Over the last two decades a large number of quality assessment instruments and checklists have focused on specific aspects of quality in the context of specific types of research (e.g. the Cochrane Risk of Bias (RoB) tool for randomized trials [16], the tool for Quality Assessment of Diagnostic Studies (QUADAS-2) [17], or the Risk Of Bias In Non-randomized Studies tool (ROBINS-I) [18]). Other instruments have addressed the reporting of results from specific study types (e.g. CONsolidated Standards for Reporting Trials (CONSORT) [19], STrengthening the Reporting of OBservational studies in Epidemiology (STROBE) [20], or Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [21]) and accordingly the reporting of protocols (e.g. Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT) [22]). The Grading, Recommendation, Assessment, Development, and Evaluation (GRADE) initiative addresses risk of bias and, in addition, imprecision, inconsistency, and indirectness as domains to assess the overall quality of a "body of evidence" for the development of evidence-based clinical guidelines [23]. These instruments and checklists are useful means to address specific aspects of quality but do not consider the research process itself.

In other research fields, including higher education [24], legal sciences [25], or political sciences [26] , the assessment of overall research quality has been described as complex, ambiguous, and a “major issue”. Increasingly, efforts have been directed towards the development of comprehensive quality frameworks [27]. Such broader approaches to quality assessment should consider the extent to which research meets the needs and expectations of stakeholders, and therefore depends on their perspective. However, the stakeholders in clinical research are numerous and their particular interests and priorities differ. Measurements of quality of clinical research may therefore be limited, or distorted, if prior consensus on a definition of quality has not been reached, and if the complexity of clinical research itself and the variety of stakeholders involved has not been taken into account. Avedis Donabedian, a pioneer in the assessment of the quality of care, declared in 1980: “What is missing (…) is a unifying theory of the definition and measurement of quality of care” (…) Before we attempt to assess the quality of care, either in general terms or in any particular site or situation, it is necessary to come to an agreement on what the elements that constitute it are” [28, 29].

This study aims to provide an overview of the existing definitions, concepts, and criteria of clinical research quality and to examine their variability by systematically synthesizing qualitative sources from the involved stakeholder groups and the medical literature. Clinical research in this context is defined as research conducted with patients to answer therapeutic, preventive, diagnostic, or prognostic questions or investigations of the mechanisms of human disease. We explicitly exclude research focusing on health care system processes, structures or policies (such as health services research or health technology assessments) and research with healthy volunteers. The findings of this study will inform the next step, i.e. the composition and structure of a comprehensive framework for clinical research quality as a common goal to increase value and reduce waste.

Methods

We conducted two systematic searches for definitions, concepts, and criteria of clinical research quality (see Box 1 for definitions of terms). We searched (i) websites and any linked documents of stakeholders in clinical research, and (ii) the published medical literature.

Search of stakeholder websites

Stakeholder website selection

We searched stakeholder organizations (national ministry of health, regulatory body, pharmaceutical industry association, academic research organization, ethics committee, patients' organization, funding agency, and initiative for clinical research) in 13 countries (Australia, Austria, Canada, France, Germany, Italy, Japan, Norway, Spain, Sweden, Switzerland, USA, UK) to provide perspectives from developed nations in different geographic regions. To identify at least one representative national stakeholder organization per stakeholder category in each of the 12 countries, we used personal contacts to one recognized expert in clinical research or public health per country. For the two contacts that did not respond (Australia, Norway), we identified the national organizations for all categories through a web search. We additionally searched for websites of inter- or supranational bodies involved in clinical research (e.g. ICH, WHO, Horizon2020, international associations) and the global 2013 Top10 pharmaceutical companies (IMSExecutive) and Contract Research Organizations (pharma-iq.com). We eventually identified publicly available websites of 155 organizations using the Google Search Engine (see S1 Table for the full list of screened organizations).

Eligibility criteria and search process

We systematically and in duplicate screened each website for a statement on a definition or concept of quality by the respective organization (e.g. "our trials are of high quality because they matter to patients", or "quality means relevant, valid, and ethical trials") using the keywords "quality" or "good" and "clinical research" or "clinical studies" or "clinical trials" or "research" in the website's search function. If we did not find a statement on quality, we extended the search to related website content, e.g. "our policy", "what we do", "standards & quality assurance" etc., as well as organizational statements, guidelines, and reports. Within these documents we repeated the search for the above search terms using the respective search function. If no statements were found through the search function, the text was manually searched for paragraphs that described either a) the standards according to which the organization performed clinical research (i.e. ICH-GCP, Declaration of Helsinki, etc.), b) criteria according to which the organization assesses the quality of clinical research (e.g. evaluation criteria of funding programs), c) the processes used to assure the quality of clinical research within an organization (e.g. "quality assurance procedures"), or d) criteria which a "good study" should fulfil within the organization. We did not consider any statements that focused on animal research, quality of life, or quality of health care without providing any definition related to clinical research. For websites presented in languages other than English

or German, text passages were translated by members of the investigative team (BvN, CPM, MMB, MR).

Search of the literature (MEDLINE)

With the help of an experienced research librarian (NB) we designed a comprehensive search strategy using MeSH terms and text words (see S1 Text for full search strategy) and conducted a systematic literature search in MEDLINE using the Ovid interface from database inception to February 27, 2015. We did not impose any language restrictions.

Eligibility criteria and selection process

We included any article describing a definition, a concept, criteria, or a checklist, guideline, or measurement instrument of quality spanning more than one quality dimension of clinical research in general or within a specific clinical discipline. We excluded any articles not suggesting a definition, concept, or criteria of clinical research quality (e.g. exclusively discussing the implementation or validation of individual quality criteria or guidelines without providing any definition related to clinical research), systematic reviews applying an assessment tool of a specific aspect of quality (e.g. systematic reviews on the reporting quality of trials in a specific field applying CONSORT [19], or articles suggesting a measurement instrument/assessment tool of one specific aspect of quality (e.g. the Jadad Scale [30]). In addition, we excluded articles that focused on animal research, quality of life, or quality of health care without providing any definition related to clinical research.

Working in pairs, methodologically trained reviewers applied the pre-defined eligibility criteria independently after undergoing a calibration process. The reviewers first screened titles and abstracts. If titles and abstracts suggested an article meeting the above mentioned inclusion criteria or if eligibility remained unclear, we obtained corresponding full texts. Disagreements were resolved by discussion and consensus.

Data extraction

We designed standardized extraction sheets suitable for qualitative data extraction (S2 Table) accompanied by an instruction manual. Before starting data extraction, the data extraction forms were piloted and teams of reviewers conducted calibration exercises to ensure consistency. We extracted text sections on the definition, concept, or criteria of quality from both literature and internet sources independently and in duplicate. Data synthesis of included articles involved categorization by overall topic, author, year of publication, article citation index (as retrieved in ISI Web of Science by 11 January 2016),

and journal name. Internet sources were categorized by stakeholder group, country, and name of organization.

Data analysis

We performed descriptive and qualitative explanatory analyses following the Framework Method [31] on definitions, concepts, and criteria of clinical research quality stratified by stakeholders and on evaluation criteria of funding agencies for clinical studies. The Framework Method belongs to a family of qualitative approaches termed thematic or content analysis, which identify commonalities and differences in qualitative data, and eventually seek to draw descriptive and/or explanatory conclusions clustered around themes. Its defining feature is the matrix output, i.e. rows (cases), columns (codes) providing a structure into which the researcher can systematically reduce the data in order to analyze it [32]. We therefore applied codes to excerpts of raw data and added or modified as new responses emerged. Codes were then grouped into clusters around similar and interrelated ideas to identify common and stakeholder-specific quality themes in an iterative process until consensus between the three investigators (BvN, MB, CPM) was reached. Themes were named after the most frequently recurring terms within the same clusters (e.g. generalizability, relevance, high quality data etc.) and were not created or imposed by the investigators.

Box 1. Glossary of working definitions, in alphabetical order

Clinical Research

Interventional and observational research addressing health care issues and involving human participants.

Concept of quality

An *implicit statement* on what clinical research quality means and comprises, e.g. which criteria are needed to ensure good quality research (often operational , e.g. "at our institution, the factors required to ensure quality are…"), or a discussion of one or multiple quality dimensions in the context of clinical research (e.g. internal validity, external validity, transparency, etc.)

Definition of quality

An *explicit statement* on what clinical research quality means and comprises, e.g. "quality of clinical research may be defined as the internal validity of study results and their applicability to patient treatment", "quality of clinical research is commonly defined as…", or "we define quality as…". May include one or multiple quality dimensions.

Quality criteria

Aspects that are described as integral part(s) of quality, e.g. adherence to guidelines, use of standard operating procedures, etc.

Quality dimension

Overarching categories of quality criteria, e.g. internal validity, external validity, relevance, transparency, etc.

Quality framework

Theoretical foundation for a definition or concept of quality spanning multiple dimensions and study phases in a matrix structure; and serving the development of quality indicators for operationalization.

Quality indicator

An instrument to assess or measure an individual quality criterion, a group of quality criteria, or a quality dimension, i.e. the operationalization of quality criteria or dimensions (e.g. how to assess the adherence to guidelines).

Quality theme

A recurrent topic in the qualitative analysis of text material about quality definitions, concepts, or criteria extracted from stakeholder websites or articles published in the literature.

Results

Definitions or concepts of clinical research quality in different stakeholder groups

We screened publicly available websites and linked documents of 155 stakeholders. Concepts of how to assure quality of clinical research or quality criteria were commonly reported among most stakeholder groups (66.4% (103/155); i.e. in 86.1% (31/36) of pharmaceutical companies or contract research organizations (CROs), 72% (18/25) of academic research organizations or initiatives, 63.6% (14/22) of international and governmental organizations, 61.9% (13/21) of regulatory agencies, 57.9% (11/19) of ethics committees, and 63.2% (12/19) of funding agencies, respectively), but this was relatively uncommon for patient organizations (31%; 4/13). However, only 12 of 155 (7.7%) institutions provided an explicit definition of the term 'clinical research quality' (pharmaceutical companies or CROs: 3/36; academic research organizations and initiatives: 3/25; international and governmental organizations: 3/22; regulatory agencies: 2/21; patient organizations: 1/13; ethics committees and funding agencies: 0/38) (S3 Table).

Qualitative analysis of the 12 definitions and the 103 quality concepts or criteria resulted in both common and stakeholder-specific quality themes often focusing on different stages of clinical research (planning/feasibility, conduct, dissemination; Table 1). Common quality themes amongst stakeholder groups included the adherence to all applicable national and international laws and regulations (e.g. ICH GCP), a scientific and methodologically rigorous approach allowing for an efficient and effective answer to the research question, credible and high quality data, the inclusion of trained study personnel, and the presence of Standard Operating Procedures (SOPs) and monitoring.

Table 1. Qualitative analysis of common and stakeholder-specific quality themes in the context of clinical research.

Stakeholder	Quality theme	Content / Explanation
All stakeholders	**Adherence to regulations & laws**	• Trial performed, data generated, documented, recorded, reported in compliance with Declaration of Helsinki, ICH-GCP & all national and international applicable regulatory requirements • Protection and respect for subject's welfare, dignity and rights in accordance with Declaration of Helsinki
	Scientific and methodological aspects of research	• Methodologically «sound» study and scientifically valid, effective & efficient answer to a scientific question • Generation of credible and high quality data
	Further common themes	• Qualified/trained personnel • Presence of Standard Operating Procedures & adequate monitoring procedures
Governmental bodies	**Relevant, transparent, & ethical research**	• Ability of a product, process, or service to satisfy stated or implied needs • Public access to information and findings • Impact on research community • Integrity, preventing poor performance and misconduct
Regulatory agencies	**Adherence to guidelines**	• Quality of evidence sufficient to support good decision making
Academic research / Clinical Trial Units / Initiatives / Networks	**Absence of bias, relevance &transparency**	• Understanding of existing evidence, assumptions explicit and justified • Particular focus on bias prevention, internal & external validity, methodological strength • Advance knowledge, bear on policy issues, address needs of patients early • The study should be compelling, useful, and relevant to stakeholders and decision makers • The study should be objective, independent, and balanced • Accurate reporting and transparency
Pharmaceutical industry/ Contract Research Organizations	**High quality data**	• Fitness for purpose / use data • Relevant to patients, HC professionals & society • Publication of all scientifically and clinically relevant information
Ethics committees / Institutional Review Boards	**Risk/benefit ratio & subject protection**	• Value enhancement of health or knowledge & benefit to community • Favorable risk/benefit ratio • Honesty, integrity, fair subject selection, free informed consent • Acknowledgement of roles of others in research • Responsible communication to the public
Patient organizations	**Patient involvement & applicability**	• Feasible and practical trials, early patient involvement • Patient-centeredness as to study procedures, inclusion/exclusion criteria and outcomes, impact on patient care • Fair subject selection & Meaningful Informed Consent • Access to quality information, during and after trial • Access to treatment after trial • Prevent risks and errors that truly matter to patient safety and the validity of the trial data
Funding agencies	**Feasibility, generalizability, & objectivity**	• Overall feasibility, no duplication of research • Important outcome to end user / potential clinical application • Evidence on comparative effectiveness & cost • Transparency / Reporting / Access to data • Inter-/ multidisciplinarity • No conflict of interest (financial/intellectual) • Internationally competitive and reproducible capacity to attract resources

Stakeholder-specific emphasis on quality themes ranged from “high quality data” (pharmaceutical industry and CROs); “adherence to guidelines” (regulatory agencies); “patient involvement and applicability of research” (patient organizations); “absence of bias, relevance, and transparency” (academic research and/or initiatives); to “feasibility, generalizability, and objectivity of research” (funding agencies). The terminology used by the stakeholders to describe these themes (e.g. relevance, transparency, feasibility), was no less open to definition than the overarching concept of “quality” and as well depends on the perspective of the observer. In general, priorities within stakeholder groups were similar across different countries. However, for national funding agencies we found considerable variation in quality criteria that were particularly emphasized as relevant for funding decisions across countries (S4 Table).

Definitions or concepts of clinical research quality in the medical literature

Our systematic MEDLINE search yielded 8’289 titles and abstracts, of which we reviewed 90 articles in full text (Fig 1). We excluded 43 full text articles from detailed analysis, because they did not discuss a definition, concept, or criteria of quality (n=18), they were systematic overviews/summaries of existing quality assessment checklists, instruments, or scores, with or without critical discussions of their validity and/or reliability (n=5), or they discussed specific measurement instruments of a single dimension of quality (n=20).

We included the remaining 47 articles for more detailed analysis (S2 Text). These provided concepts on how to assure or improve overall clinical research quality in specific contexts (e.g. at an academic institution, in a specific country, in a specific industry setting, or in a specific medical field; n=18), or how to improve quality assessment (e.g. of RCTs, in radiology or hepatology research; n=6). Measurement instruments or checklists that spanned more than one quality dimension were reported in 23 articles. A large proportion of these tools provided indicators on how to assess bias (n=21). Almost half of them covered indicators on precision (n=16), external validity (n=16), or reporting quality (n=14). Some tools additionally covered innovation aspects (n=8) or ethical considerations (n=4). None of the reviewed articles provided a definition or concept of clinical research quality spanning the encountered range of quality dimensions reflected by stakeholder perspectives such as ethical conduct, patient safety, patient values and preferences, absence of bias, precision, external validity, relevance, generalizability, transparency, infrastructure, and sustainability. Furthermore, we could not identify a definition or concept simultaneously covering several dimensions and differentiating between consecutive stages of research (e.g. study planning, conduct and dissemination), independent of a specific medical field or study setting.

Overall, we identified eight (8.9%) of 90 articles that provided an explicit definition of the term 'clinical research quality' (Table 2). The definitions therein span quality from methodological dimensions such as internal validity, external validity, or precision, and operational criteria including adherence to guidelines and applicable regulations (ethical conduct), to the effect of research at the societal level (relevance). None of the definitions appeared to be comprehensive either in terms of quality dimensions, research stages, or stakeholder perspectives. Five of the eight articles were cited less than 10 times in ISI Web of Science™ by 11.01.2016 (Table 2).

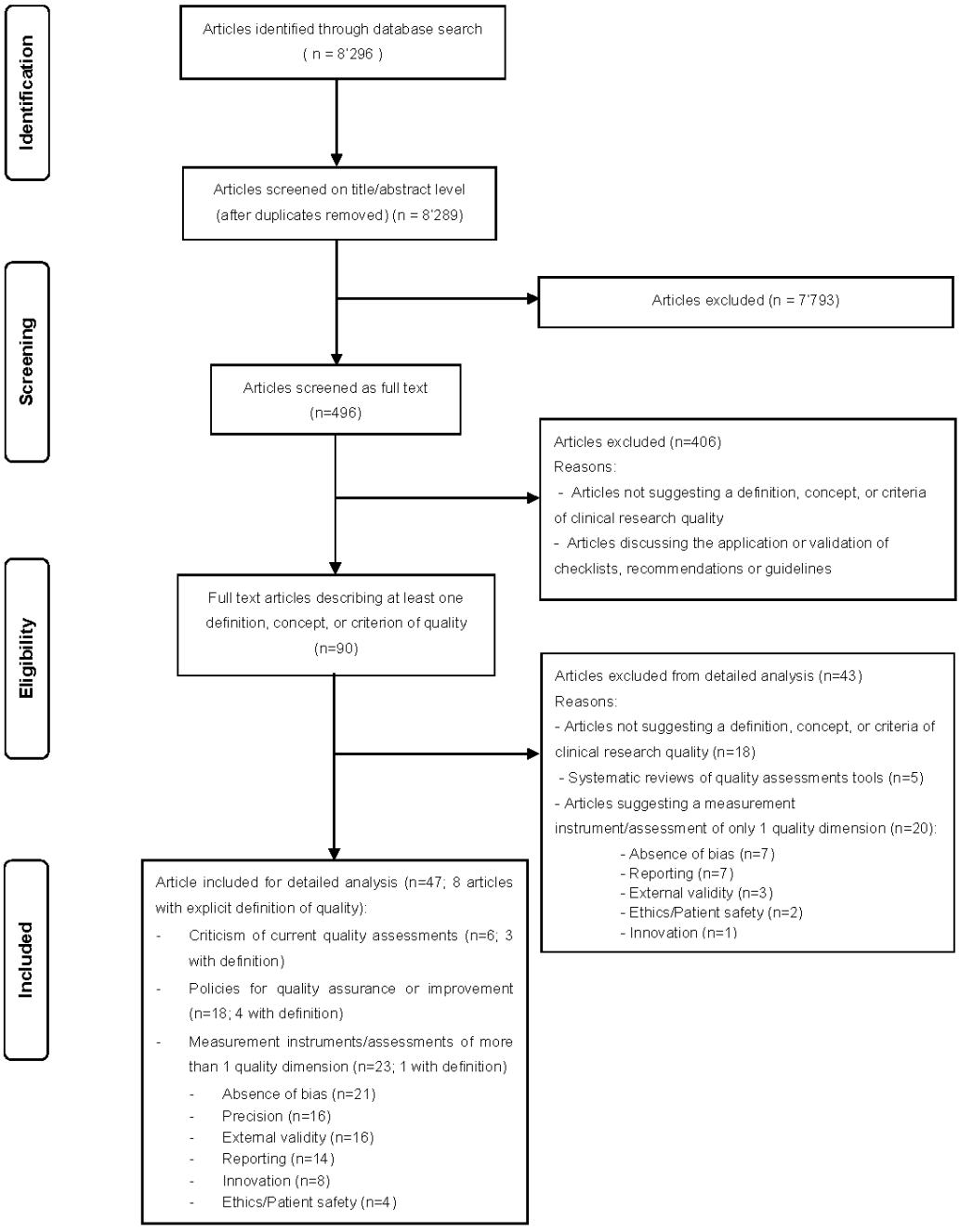


Fig 1. Article flow diagram

Table 2. Characteristics of articles providing an explicit definition of clinical research quality; by author (n=8).

Author(s), Year	Title	Journal	Setting	Quality definition	Cit.[a]
Moher, Jadad et. al. (1996) [33]	Assessing the quality of randomized controlled trials. Current issues and future directions.	International Journal of Technology Assessessment in Health Care	RCTs	(…) Quality is a construct (concept) that can be defined in many ways, including the literary aspects for the report of a trial or its external validity, i.e. the degree to which it is possible to generalize trial results. Our focus on one important aspect of methodologic quality (hereafter simply "Quality"), internal validity, which we define as the "confidence that the trial design, conduct, analysis, and presentation has minimized or avoided biases in its Intervention comparisons." However, we recognize that this definition excludes other methodologic aspects of quality, for example, those concerned with the precision and reliability of measurements or estimation of compliance. (…)	244
Verhagen, de Vet et al. (2001) [34]	The art of quality assessment of RCTs included in systematic reviews	Journal of Clinical Epidemiology	Systematic review of RCTs	(…) Quality of RCTs has recently been defined as: "the likelihood of the trial design to generate unbiased results". This definition covers only the dimension of internal validity. During the development of the "Delphi list" for quality assessment, the participants, all experts in the field of RCTs, failed to reach consensus on a specific definition, but did agree that the concept of quality should comprise more than internal validity alone. From this context we propose the following definition of quality: the likelihood of the trial design to generate unbiased results, that are sufficiently precise and allow application in clinical practice. (…)	125
Njie and Thomas (2001) [35]	Quality issues in clinical research and the implications on health policy (QICRHP)	Journal of Professional Nursing	General	(…)In this article, quality in clinical research is the process of developing and implementing guidelines to ensure the inclusion of all pertinent aspects of the research process, ensure accountability of research team members, adherence to protocol guidelines, and maintenance of study integrity and merit. (…)	1
Franck, Pendleton et al. (2004) [36]	Quality assurance for clinical research: challenges in implementing research governance in UK hospitals	International Journal of Health Care Quality Assurance Incorporating Leadership in Health Services	UK hospitals	(…) The essential elements of high quality research conduct derived from this body of literature are: research ethics (dignity, rights, safety, well-being of research participants); scientific quality, adherence to regulations (health and safety, medicines and devices); and information integrity (data protection, dissemination, financial and intellectual property). (…)	2
Switula (2006) [37]	The concept of quality in clinical research	Science & Engineering Ethics	General	(…) Quality in clinical research may be defined as compliance with requirements together with credibility and reliability of the data obtained. In the spirit of ISO, we may define quality in the clinical research process pictured above as the positive characteristics of the end product, that is the reliability and credibility of information collected during the clinical research process. Quality of research also means compliance of the whole trial process with pre-defined requirements. The customers of the clinical research define these requirements. (…)	3
Krestin (2008) [38]	Evaluating the Quality of Radiology Research: What Are the Rules of the Game?	Radiology	Radiology	(…)"I believe that research quality can be defined as the contribution of research to national and global social, economic, and scientific progress—that is, the effect of research at the societal level contribution of research to society." (…)	1
Bhatt (2011) [39]	Quality of clinical trials: A moving target	Perspectives in Clinical Research	FDA	(…) Quality of clinical trials depends on data integrity and subject protection. (…)	8
Balshem, Helfand et al. (2011) [40]	GRADE guidelines: 3. Rating the quality of evidence.	Journal of Clinical Epidemiology	Quality of Evidence	(…)''Quality'' as used in GRADE means more than risk of bias and so may also be compromised by imprecision, inconsistency, indirectness of study results, and publication bias. In addition, several factors can increase our confidence in an estimate of effect. GRADE provides a systematic approach for considering and reporting each of these factors. (…)	690

[a] Citations in Web of Science, last updated 11.01.2016 Abbreviations: FDA, US Food and Drug Administration; RCT, Randomized Controlled Trial; UK, United Kingdom

Discussion

Summary of findings

Our systematic review of stakeholder websites and the medical literature showed that quality of clinical research is frequently discussed, but rarely defined. Although stakeholder groups seem to agree on a basic concept of quality, their emphasis in the conceptualization of clinical research quality varies widely. The medical literature contains many articles discussing approaches to measurement or assessments of quality without prior definition of the term itself, and without reflecting the diversity of stakeholder needs, interests and expectations. A major proportion of these identified quality assessments aim to evaluate the "methodological rigor of randomized controlled trials". The definition of "methodological rigor" in itself, however, varies substantially between the reported tools. Most authors suggested assessing methodological quality based on the presence or absence of measures to prevent bias. Others included dimensions such as external validity, reporting, or relevance of the study in question. We did not, however, identify a definition or concept including multiple dimensions or differentiating between consecutive stages of research across medical fields, or study settings. Although a comprehensive "definition" of quality may be difficult, a "concept" or "framework" of research quality, rather than a "definition", could span all research stages and include more than an assessment focused on one aspect of quality. A more comprehensive approach to quality assessment, i.e. ranging from conceptualization to dissemination of a study as proposed by the authors of the Lancet series on "increasing value, reducing waste" [1, 4-7], rather than evaluation of the final published product, would assist in identification of errors that matter at earlier stages, and therefore support reduction of research "waste" more efficiently.

Strengths and limitations

To our knowledge this is the first systematic survey addressing definitions and concepts of clinical research quality. Our systematic approach was suited to detect knowledge gaps, and to examine overlap and differences in perspectives of clinical research quality across stakeholder groups. Further strengths of this study include our consideration of websites and any linked documents from a large number of stakeholders in 13 different countries in addition to a Medline search. Methodologically trained investigators screened articles and websites in duplicate following a pre-specified instruction manual and undergoing a calibration process.

We acknowledge the following limitations: Although we consider our approach comprehensive, we searched only Medline as electronic database and relied on search terms in the title, abstract or other records. Articles in journals not indexed in Medline or

providing some definition of research quality in the main text only might have been overlooked. However, Medline covers the most impactful journals and articles in current medical research and articles specifically focusing on quality of clinical research most likely mention this prominently. We may have missed definitions on websites despite screening these in duplicate. We would, however, expect stakeholder groups in clinical research to be transparent and proactive in defining such an important cornerstone of their activities, similarly to efforts in the field of clinical care quality. When coding our findings from the website search as well as from the Medline search we felt that we reached saturation, i.e. the last excerpts from websites or journal articles on aspects of clinical research quality did not bring new insights. The coding and qualitative analysis naturally involved subjective judgments of investigators, which we controlled by performing analyses in triplicate (BvN, CPM, MB) and comparing codes, findings, and interpretations until we reached consensus. Further, we acknowledge that our survey solely portrays perceptions on the quality of research in high income nations that may not necessarily overlap with those of low- or middle-income countries. With a growing percentage of clinical research being conducted in these geographies, a further study investigating quality perceptions of local stakeholders taking into consideration societal aspects and beliefs would be of importance. Finally, we did not conduct a detailed survey of experts or stakeholder groups - except for national funding agencies - nor did we conduct interviews with representatives of these groups to explore reasons for the paucity of explicit definitions.

Comparison with other studies and implications

We are not aware of any other systematic survey on definitions, concepts, or criteria of overall clinical research quality. A similar approach has been taken by other authors to develop a framework for excellence, however with a distinct focus on translational cancer research [41]. In the field of health care quality, the focus of assessments has more and more shifted from process-based measurements towards the evaluation of patient outcomes and patient satisfaction [42-44]. Clinical research conducted in this context of “patient-centered” care would explicitly warrant the engagement and involvement of patients in setting priorities. However, patients (or their representatives) are only rarely considered when discussing the quality of research that might impact their care [45-47]. In our analysis, we also found patients to be surprisingly underrepresented. First, patient organizations were among the last in providing definitions or concepts of clinical research quality. Second, only six (14%) of a total of 43 quality measurement tools or assessments covered an item on patient safety and/or rights (Fig 1). Most efforts in quality assessments so far were taken to ensure compliance with guidelines, methodologically rigorous designs and valid study results. While this may ultimately serve the treatment of disease, we were expecting the

clinical research enterprise to put unmet medical need and applicability to the patient population first. Compared with medical care, the clinical research machinery still seems to function with relatively low engagement of the end user (patients) of the product.

Furthermore, while we expected variations in the perception of quality across stakeholder groups, we were surprised how different and vaguely defined some of the concepts were. For example, an explicit definition of “high quality data” may be as dependent on the perspective of the observer as the definition of “high quality research”. It may be linked to concepts such as relevance of the data and absence of errors in the data or the way the data is collected. Similarly, the quality criteria used by funding agencies such as “impact”, “relevance”, or “feasibility” varied in their clarity and elaboration. Public funding agencies have a major role in terms of defining what and how research topics are investigated. Those who use these criteria to evaluate proposals are still left with subjective interpretation, while applicants may aim to provide the readers with these buzzwords with not much reflection on their meaning.

There remains considerable ambiguity in the use of current quality criteria across and within stakeholder groups. Unless carefully explained, these concepts can be easily misinterpreted by the stakeholders. Finding consensus on a common definition or concept of clinical research quality across national borders, stakeholder groups, and study types may therefore seem arduous; assessments of methodological quality do not, however, suffice. Existing quality guidelines such as ICH GCP have not been developed based on consensus across the full range of stakeholder groups, but only between regulatory experts and industry [15]. Existing quality assessment tools predominantly cover single aspects of quality, or particular research stages. Furthermore, there is a lack of approaches tailored to stakeholder requirements in assessing the quality of clinical research, e.g. from a patient’s perspective on how to choose a “good trial”, or from a funding agency’s perspective on how to assess the quality of studies before, during, and after the funding period. The authors of a follow-up study to the 2014 Lancet series reported that academic institutions in particular had paid only little attention to their recommendations on how to increase value in research. Practical guidance on how to implement these recommendations is so far lacking and urgently needed to increase value of academic research at all stages.

Conclusions

This systematic survey serves as a first step of evidence summary to inform the development of a comprehensive framework of clinical research quality. It showed that definitions of clinical research quality are rarely provided and the existing definitions fall short of a theoretical or empirical framework across different study designs and stages and considering the variety of stakeholders involved. Based on our findings, a practically applicable framework needs to include the encountered quality dimensions such as ethical conduct, patient safety/rights/priorities, internal validity, precision of results, generalizability or external validity, scientific and societal relevance, transparency and accessibility of information, research infrastructure and sustainability) and consider different study stages such as planning, conduct, and dissemination. We plan to circulate framework drafts amongst stakeholder representatives of all eight groups until consensus on structure and content is reached, and to operationalize the framework through the development of instruments guiding stakeholder groups (e.g. academic institutions or funding agencies) in the comprehensive quality assessment of the full clinical research continuum.

Acknowledgements

We would like to thank Dr. Priya Satalkar for her guidance and support with the qualitative analysis.

References

[1] Al-Shahi Salman R, Beller E, Kagan J, Hemminki E, Phillips RS, Savulescu J, et al. Increasing value and reducing waste in biomedical research regulation and management. Lancet. 2014;383:176-85.
[2] Kasenda B, von Elm E, You J, Blumle A, Tomonaga Y, Saccilotto R, et al. Prevalence, characteristics, and publication of discontinued randomized trials. Jama. 2014;311:1045-51.
[3] Chalmers I, Bracken MB, Djulbegovic B, Garattini S, Grant J, Gulmezoglu AM, et al. How to increase value and reduce waste when research priorities are set. Lancet. 2014;383:156-65.
[4] Chan AW, Song F, Vickers A, Jefferson T, Dickersin K, Gotzsche PC, et al. Increasing value and reducing waste: addressing inaccessible research. Lancet. 2014;383:257-66.
[5] Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, Julious S, et al. Reducing waste from incomplete or unusable reports of biomedical research. Lancet. 2014;383:267-76.
[6] Ioannidis JP, Greenland S, Hlatky MA, Khoury MJ, Macleod MR, Moher D, et al. Increasing value and reducing waste in research design, conduct, and analysis. Lancet. 2014;383:166-75.
[7] Macleod MR, Michie S, Roberts I, Dirnagl U, Chalmers I, Ioannidis JP, et al. Biomedical research: increasing value, reducing waste. Lancet. 2014;383:101-4.
[8] Ioannidis JP. Why Most Clinical Research Is Not Useful. PLoS Med. 2016;13:e1002049.
[9] Juni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. Bmj. 2001;323:42-6.
[10] The International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use, ICH Harmonised Tripartite Guideline: Guideline for Good Clinical Practice. https://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E6/E6_R1_Guideline.pdf. Accessed 11 Nov 2016.
[11] International conference on harmonisation of technical requirements for registration of pharmaceuticals for human use, Intergrated Addendum to ICH E6(R1): Guideline for Good Clinical Practice E6R(2). http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E6/E6_R2_Step_4.pdf. Accessed 11 Nov 2016.
[12] Stewart PM, Stears A, Tomlinson JW, Brown MJ. Regulation--the real threat to clinical research. Bmj. 2008;337:a1732.
[13] McMahon AD, Conway DI, Macdonald TM, McInnes GT. The unintended consequences of clinical trials regulations. PLoS Med. 2009;3:e1000131.
[14] Woolf SH. Practice guidelines, a new reality in medicine. II. Methods of developing guidelines. Arch Intern Med. 1992;152:946-52.
[15] Grimes DA, Hubacher D, Nanda K, Schulz KF, Moher D, Altman DG. The Good Clinical Practice guideline: a bronze standard for clinical research. Lancet. 2005;366:172-4.
[16] Higgins JP, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. Bmj. 2011;343:d5928.
[17] Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155:529-36.
[18] Sterne JA, Hernan MA, Reeves BC, Savovic J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. Bmj. 2016;355:i4919.
[19] Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. Ann Intern Med. 2001;134:663-94.
[20] Vandenbroucke JP, von Elm E, Altman DG, Gotzsche PC, Mulrow CD, Pocock SJ, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. Int J Surg. 2014;12:1500-24.

[21] Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. Ann Intern Med. 2009;151:W65-94.
[22] Chan AW, Tetzlaff JM, Gotzsche PC, Altman DG, Mann H, Berlin JA, et al. SPIRIT 2013 explanation and elaboration: guidance for protocols of clinical trials. Bmj. 2013;346:e7586.
[23] Guyatt GH, Oxman AD, Kunz R, Vist GE, Falck-Ytter Y, Schunemann HJ. What is "quality of evidence" and why is it important to clinicians? Bmj. 2008;336:995-8.
[24] Patrick WJ, Stanley EC. Assessment of Research Quality. Research in Higher Education. 1996;37:23-42.
[25] Castermans AG. Evaluating Dutch Legal Research Quality Assessment. Evaluating Dutch Legal Research Quality Assessment. 2015:447-73.
[26] Kellow A. Assessing Political Science Quality: 'Excellence in Research for Australia'. European Political Science. 2012;11:567-80.
[27] Cabral AP, Huet I. Assessment of Research Quality in Higher Education: Contribution for an Institutional Framework. Procedia - Social and Behavioral Sciences. 2014;116:1528-32.
[28] Donabedian A. The Definition of Quality and Approaches to Its Assessment. Ann Arbor: Health Administration Press. 1980;Volume I: Explorations in Quality Assessment and Monitoring 176 pp.
[29] Donabedian A. The quality of care. How can it be assessed? Jama. 1988;260:1743-8.
[30] Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin Trials. 1996;17:1-12.
[31] Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC Med Res Methodol. 2013;13:1-8.
[32] Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC Med Res Methodol. 2013;13:117.
[33] Moher D, Jadad AR, Tugwell P. Assessing the quality of randomized controlled trials. Current issues and future directions. Int J Technol Assess Health Care. 1996;12:195-208.
[34] Verhagen AP, de Vet HC, de Bie RA, Boers M, van den Brandt PA. The art of quality assessment of RCTs included in systematic reviews. J Clin Epidemiol. 2001;54:651-4.
[35] Njie VP, Thomas AC. Quality issues in clinical research and the implications on health policy (QICRHP). J Prof Nurs. 2001;17:233-42.
[36] Franck LS, Pendleton E, Pittam B, Preece M, Aynsley-Green A. Quality assurance for clinical research: challenges in implementing research governance in UK hospitals. Int J Health Care Qual Assur Inc Leadersh Health Serv. 2004;17:239-47.
[37] Switula D. The concept of quality in clinical research. Sci Eng Ethics. 2006;12:147-56.
[38] Krestin GP. Evaluating the Quality of Radiology Research: What Are the Rules of the Game? Radiology. 2008;249:418-24.
[39] Bhatt A. Quality of clinical trials: A moving target. Perspect Clin Res. 2011;2:124-8.
[40] Balshem H, Helfand M, Schunemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol. 2011;64:401-6.
[41] Rajan A, Caldas C, van Luenen H, Saghatchian M, van Harten WH. Assessing excellence in translational cancer research: a consensus based framework. J Transl Med. 2013;11:274.
[42] Jenkinson C, Coulter A, Bruster S. The Picker Patient Experience Questionnaire: development and validation using data from in-patient surveys in five countries. Int J Qual Health Care. 2002;14:353-8.
[43] Urden LD. Patient satisfaction measurement: current issues and implications. Lippincotts Case Manag. 2002;7:194-200.
[44] Dawn AG, Lee PP. Patient satisfaction instruments used at academic medical centers: results of a survey. Am J Med Qual. 2003;18:265-9.

[45] Dugas M, Trottier M, Chipenda Dansokho S, Vaisson G, Provencher T, Colquhoun H, et al. Involving members of vulnerable populations in the development of patient decision aids: a mixed methods sequential explanatory study. BMC Med Inform Decis Mak. 2017;17.
[46] Coulter A, Stilwell D, Kryworuchko J, Mullen PD, Ng CJ, van der Weijden T. A systematic development process for patient decision aids. BMC Med Inform Decis Mak. 2013;13:S2.
[47] Frank L, Forsythe L, Ellis L, Schrandt S, Sheridan S, Gerson J, et al. Conceptual and practical foundations of patient engagement in research at the patient-centered outcomes research institute. Qual Life Res. 2015;24:1033-41.

Supporting information

S1 Text. Literature search strategy

1 Quality Control/
2 total quality management/
3 Quality Improvement/
4 Quality Indicators, Health Care/
5 exp *Reproducibility of Results/
6 good clinical practic*.mp.
7 or/1-6
8 biomedical research/ or clinical nursing research/
9 (biomedical and research).ti,ab.
10 (clinical adj2 research).ti,ab.
11 Randomized Controlled Trial* as Topic.mp. or Randomized Controlled Trials as Topic/
12 clinical trial* as topic.mp. or Clinical Trials as Topic/
13 Research/st [Standards]
14 or/8-13
15 7 and 14
16 ((quality or valid* or data integrity) adj4 (clinical stud* or clinical data or clinical trial* or randomized trial* or randomised trial* or control* trial*)).ti,ab.
17 ((quality or valid* or data integrity) adj4 (clinical or medical or human* or patient*) adj2 research*).ti,ab.
18 16 or 17
19 15 or 18
20 or/1-6
21 14 and 20
22 18 or 21
23 (approach or assessment* or assessing or assurance or checklist* or check list*).tw.
24 (code of conduct or concepts or concept or clinimetric* or definition* or evaluation*).tw.
25 (framework* or guideline* or guidance or instrument or instruments or indicators or indicator).tw.
26 (predictor* or measurement* or measures or measure or process or rating*).tw.
27 (scale? or score? or standard? or norm? or system? or tool? or dimension? or item?).tw.
28 (factor? or criteria or principle? or grading or grade or attributes or metrics or monitor).tw.
29 (recommendation? or priority or priorities or construct? or determinant? or project?).tw.
30 (report? or point? or categories or category or summaries or summary or ranking?).tw.
31 (statistic? or term? or feature? or characteristic? or profile? or pattern? or rule? or idea).tw.
32 (theory or paradigm? or consensus or statement or promotion or promoting or test?).tw.
33 (improvement* or improving or increasing).tw.
34 or/23-33
35 22 and 34

S2 Table. Data extraction forms

A) Excel based extraction form for website search

Organisation name	Location	Type of statement source	Type of quality statement	Text/Statement	Comments	Reference	Link
	International	Government Document	Definition				
	USA	Journal Article	Discussion				
	EU	Legal Rule or Regulation	Operationalisation				
	Australia	Magazine Article	Other				
	Canada	Personal Communication					
	France	Press Release					
	Germany	Report					
	Italy	Statute					
	Japan	Web Page					
	Norway	Guideline					
	UK	Other					
	Switzerland						
	Spain						
	Sweden						
	Austria						
	Other						

B) Web-based full text extraction form for literature search

Name	Type	Label	Options
relevance	list	The article is relevant (according to inclusion/exclusion criteria, for details see "info")	yes
			no, exclude
relevance	text	Why is article not relevant? Please comment.	
article_type	list	The article's main focus is on	Definition of quality
			Assurance/assessment of quality
data	list	Article results are based on	Expert consensus
			Author opinion
			Empirical data
			Other
keytext	text	Copy key statement on quality definition from article	
keyword	text	Quality Item/s (separated by ;)	
dimension	list	To which quality dimension would you add the item/s?	Absence of bias
			Precision
			External validity
			Innovation/Relevance
			Reporting/Transparency
			Education/Training
			GCP/Patient safety
			Other/New domain
scale	list	Was an existing quality scale/checklist/score used?	yes
			no
namescale	text	Please provide the name of the scale/checklist/score used	
dimchecklist	list	Which quality dimension/s does the checklist/scale assess?	Absence of bias
			Precision
			External validity
			Innovation/Relevance
			Reporting/Transparency
			Education/Training
			GCP/Patient safety
			Other/New domain
indicators	list	Are additional indicators/metrics/measurement instruments or assessment methods of quality in clinical research mentioned?	yes
			no
nameindicator	text	Please state what these indicators/metrics/instruments or assessments are:	
addind	list	For which quality dimension are these indicators/metrics/assessments	Absence of bias

		representative?	Precision
			External validity
			Innovation/Relevance
			Reporting/Transparency
			Education/Training
			GCP/Patient Safety
			Other/new domain
comments	text	Comments	

S2 Text. References of eligible articles

[1] Berger VW, Alperson SY. A general framework for the evaluation of clinical trial quality. Rev Recent Clin Trials. 2009;4:79-88.
[2] Berk PD, Sacks HS. Assessing the quality of randomized controlled trials: quality of design is not the only relevant variable. Hepatology. 1999;30:1332-4.
[3] Verhagen AP, de Vet HC, de Bie RA, Boers M, van den Brandt PA. The art of quality assessment of RCTs included in systematic reviews. J Clin Epidemiol. 2001;54:651-4.
[4] Krestin GP. Evaluating the Quality of Radiology Research: What Are the Rules of the Game? Radiology. 2008;249:418-24.
[5] Dechartres A, Charles P, Hopewell S, Ravaud P, Altman DG. Reviews assessing the quality or the reporting of randomized controlled trials are increasing over time but raised questions about how quality is assessed. J Clin Epidemiol. 2011;64:136-44.
[6] Moher D, Jadad AR, Tugwell P. Assessing the quality of randomized controlled trials. Current issues and future directions. Int J Technol Assess Health Care. 1996;12:195-208.
[7] Clark GT, Mulligan R. Fifteen common mistakes encountered in clinical research. J Prosthodont Res. 2011;55:1-6.
[8] Kleppinger CF, Ball LK. Building quality in clinical trials with use of a quality systems approach. Clin Infect Dis. 2010;51 Suppl 1:S111-6.
[9] Rosenblum D, Alving B. The role of the clinical and translational science awards program in improving the quality and efficiency of clinical research. Chest. 2011;140:764-7.
[10] Weihrauch TR, Baumbauer E. [Drug research in Germany--problems and chances--from the viewpoint of the drug industry]. Arzneimittelforschung. 1998;48:1047-50.
[11] Beger HG, Schwarz A. Clinical research in surgery: questions but few answers. Langenbecks Arch Surg. 1998;383:300-5.
[12] Prescott RJ, Counsell CE, Gillespie WJ, Grant AM, Russell IT, Kiauka S, et al. Factors that limit the quality, number and progress of randomised controlled trials. Health Technol Assess. 1999;3:1-143.
[13] Njie VP, Thomas AC. Quality issues in clinical research and the implications on health policy (QICRHP). J Prof Nurs. 2001;17:233-42.
[14] Holland WW, Mosbech J. International quality control of national scientific standards. J Intern Med. 1993;234:1-2.
[15] Knatterud GL, Rockhold FW, George SL, Barton FB, Davis CE, Fairweather WR, et al. Guidelines for quality assurance in multicenter trials: a position paper. Control Clin Trials. 1998;19:477-93.
[16] Strause LG, Vogel JR. The clinical research triad: how can we ensure quality in out-sourced clinical trials? Qual Manag Health Care. 1999;7:23-9.
[17] American Society of Clinical O. American Society of Clinical Oncology policy statement: oversight of clinical research. J Clin Oncol. 2003;21:2377-86.
[18] Ottevanger PB, Therasse P, van de Velde C, Bernier J, van Krieken H, Grol R, et al. Quality assurance in clinical trials. Crit Rev Oncol Hematol. 2003;47:213-35.
[19] Franck LS, Pendleton E, Pittam B, Preece M, Aynsley-Green A. Quality assurance for clinical research: challenges in implementing research governance in UK hospitals. Int J Health Care Qual Assur Inc Leadersh Health Serv. 2004;17:239-47.
[20] Warden D, Rush AJ, Trivedi M, Ritz L, Stegman D, Wisniewski SR. Quality improvement methods as applied to a multicenter effectiveness trial--STAR D. Contemp Clin Trials. 2005;26:95-112.
[21] Switula D. The concept of quality in clinical research. Sci Eng Ethics. 2006;12:147-56.
[22] Zon R, Meropol NJ, Catalano RB, Schilsky RL. American Society of Clinical Oncology Statement on minimum standards and exemplary attributes of clinical trial sites. J Clin Oncol. 2008;26:2562-7.
[23] Bhatt A. Quality of clinical trials: A moving target. Perspect Clin Res. 2011;2:124-8.
[24] Hanna M, Minga A, Fao P, Borand L, Diouf A, Mben JM, et al. Development of a checklist of quality indicators for clinical trials in resource-limited countries: the French

National Agency for Research on AIDS and Viral Hepatitis (ANRS) experience. Clin. 2013;10:300-18.
[25] Powe NR, Kinnison ML, Steinberg EP. Quality assessment of randomized controlled trials of contrast media. Radiology. 1989;170:377-80.
[26] Newcastle Ottawa Scale.
[27] Chalmers TC, Smith H, Jr., Blackburn B, Silverman B, Schroeder B, Reitman D, et al. A method for assessing the quality of a randomized control trial. Control Clin Trials. 1981;2:31-49.
[28] Cho MK, Bero LA. Instruments for assessing the quality of drug studies published in the medical literature. Jama. 1994;272:101-4.
[29] Jefferson T, Wager E, Davidoff F. Measuring the quality of editorial peer review. Jama. 2002;287:2786-90.
[30] Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin Trials. 1996;17:1-12.
[31] Manchikanti L, Hirsch JA, Cohen SP, Heavner JE, Falco FJ, Diwan S, et al. Assessment of methodologic quality of randomized trials of interventional techniques: development of an interventional pain management specific instrument. Pain physician. 2014;17:E263-90.
[32] Kocsis JH, Gerber AJ, Milrod B, Roose SP, Barber J, Thase ME, et al. A new scale for assessing the quality of randomized clinical trials of psychotherapy. Compr Psychiatry. 2010;51:319-24.
[33] Nass SJ, Balogh E, Mendelsohn J. A National Cancer Clinical Trials Network: recommendations from the Institute of Medicine. Am J Ther. 2011;18:382-91.
[34] Arrive L, Renard R, Carrat F, Belkacem A, Dahan H, Le Hir P, et al. A scale of methodological quality for clinical studies of radiologic examinations. Radiology. 2000;217:69-74.
[35] Balk EM, Bonis PA, Moskowitz H, Schmid CH, Ioannidis JP, Wang C, et al. Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. Jama. 2002;287:2973-82.
[36] Barske HL, Baumhauer J. Quality of research and level of evidence in foot and ankle publications. Foot Ankle Int. 2012;33:1-6.
[37] Birch S. Clinical research on acupuncture. Part 2. Controlled clinical trials, an overview of their methods. J Altern Complement Med. 2004;10:481-98.
[38] Bizzini M, Childs JD, Piva SR, Delitto A. Systematic review of the quality of randomized controlled trials for patellofemoral pain syndrome. J Orthop Sports Phys Ther. 2003;33:4-20.
[39] Bornhoft G, Maxion-Bergemann S, Wolf U, Kienle GS, Michalsen A, Vollmar HC, et al. Checklist for the qualitative evaluation of clinical studies with particular focus on external validity and model validity. BMC Med Res Methodol. 2006;6:56.
[40] Coleman BD, Khan KM, Maffulli N, Cook JL, Wark JD. Studies of surgical outcome after patellar tendinopathy: clinical significance of methodological deficiencies and guidelines for future studies. Victorian Institute of Sport Tendon Study Group. Scand J Med Sci Sports. 2000;10:2-11.
[41] Tate RL, McDonald S, Perdices M, Togher L, Schultz R, Savage S. Rating the methodological quality of single-subject designs and n-of-1 trials: introducing the Single-Case Experimental Design (SCED) Scale. Neuropsychol. 2008;18:385-401.
[42] Yates SL, Morley S, Eccleston C, de CWAC. A scale for rating the quality of psychological trials for pain. Pain. 2005;117:314-25.
[43] Scholler K, Licht S, Tonn JC, Uhl E. Randomized controlled trials in neurosurgery--how good are we? Acta Neurochir (Wien). 2009;151:519-27; discussion 27.
[44] Sindhu F, Carpenter L, Seers K. Development of a tool to rate the quality assessment of randomized controlled trials using a Delphi technique. J Adv Nurs. 1997;25:1262-8.
[45] Manchikanti L, Hirsch JA, Heavner JE, Cohen SP, Benyamin RM, Sehgal N, et al. Development of an interventional pain management specific instrument for methodologic quality assessment of nonrandomized studies of interventional techniques. Pain physician. 2014;17:E291-317.

[46] Balshem H, Helfand M, Schunemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol. 2011;64:401-6.
[47] Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. J Epidemiol Community Health. 1998;52:377-84

S3 Table. Quality definitions found through systematic internet search; by institution. Total number of institutions screened = 155.

Organization	Geographic scope	Quality statement
Governmental bodies / Jurisdiction		
World Health Organisation (WHO)	International	(…) "Quality" is a measure of the ability of a product, process, or service to satisfy stated or implied needs. A high quality product readily meets those needs. In the context of a clinical trial, quality may apply to data (e.g., data are accurate and reliable) or processes (e.g., compliance with the study protocol and GCP; ensuring informed consent; adequate data handling and record-keeping, etc.). (See WHO GCP Principles 6: Protocol Compliance; 7: Informed Consent; 11: Records) (…) For all studies involving human subjects, even in the early stages (whether discovery or development), Good Clinical Practices are the correct quality and ethical standards. Tight national regulations ensure patient safety and methodological quality of clinical trials. (…)
Bundesamt für Gesundheit (BAG)	Switzerland	(…) **Art. 4** Scientific quality The sponsor and the investigator of a clinical trial shall ensure scientific quality. In particular: a.they shall define a research question based on the current state of scientific knowledge; b.they shall use an appropriate scientific methodology; and c.they shall ensure the availability of the resources required for the clinical trial and provide the necessary infrastructure (…)
Department of Health (DoH)	UK	(…)The key elements of a quality research culture are: • respect for participants' dignity, rights, safety and wellbeing; • valuing the diversity within society; • personal and scientific integrity; • leadership; • honesty; • accountability; • openness; • clear and supportive management. (…)
Regulatory Agencies/HTA Bodies		

Food and Drug Administration (FDA)	USA	(…) "Quality" is characterized by the ability to effectively and efficiently answer the intended question about the benefits and risks of a medical product (therapeutic or diagnostic) or procedure while ensuring protection of human subjects. (…) Elements of a quality clinical study: - Scientifically valid and ethically sound experimental design - Adequate protection of subjects rights, safety, and welfare - Qualified personnel - "Adequate" monitoring - Current, complete, and accurate data (…)
European Medicines Agency (EMA)	EU	(…) Quality in this context is commonly defined as fitness for purpose. Clinical research is about generating information to support decision making while protecting the safety and rights of participating subjects. The quality of information generated should therefore be sufficient to support good decision making. (…)
Pharmaceutical Manufacturers & Contract Research Organizations		
AstraZeneca	UK	(…) Quality in clinical research may be defined as… • Reliability and credibility of information providing an answer to a scientific question • Compliance of the trial process with defined requirements (Nach ISO 9000: A quality is a set of characteristics that a product or service must have to satisfy needs and expectations of the customer. • Product of clinical research process: information. • Customers of clinical research : Society, Research subjects, sponsors, regulatory authorities, hospitals/institutions, IECs (…)
Pfizer	USA	(…) Components for Quality: Clinical research quality is designed and embedded in the clinical trial processes and study protocol well in advance of enrollment of the first patient. Components of the quality process related to clinical trial sites include: • Creating, implementing, and upholding standard operating procedures (SOPs) for trial execution • A quality scientific and medical design of the protocol • Clinical investigator and site pre-assessment and selection • Regulatory agency and ethics committee approval • Developing and providing appropriate informed consent (language, transparency of benefits and risks) and obtaining ethics committee approval of the informed consent process • Investigator meetings and training • Adequate recording and reporting of data • Periodic monitoring • Audits
Target Health Inc.	USA	(…) A "quality clinical trial" is one where 1) there is "absence of errors that matter" and 2) "are the data fit for use/purpose." Errors "that matter" are those that have a 1) meaningful impact on patient safety and/or 2) Interpretation of trial results. (…)

Clinical Research Initiatives / Academic Clinical Research Organizations		
COCHRANE Collaboration	International	(…) Quality : A vague notion of the methodological strength of a study, usually indicating the extent of bias prevention.(…)
DEPLHI	International	(…) Quality is a set of parameters in the design and conduct of a study that reflects the validity of the outcome, related to the external and internal validity and the statistical model used. (…)
Swiss Group for Clinical Cancer Research (SAKK)	Switzerland	(…) Quality is defined by several aspects in our organization. In general quality means the evaluation if we meet specific requirements in the development and conduct of our trials. These requirements are defined on different levels: a) The law (HRA): local applicable law to conduct clinical research b) In international guidelines (ICH GCP, GMP Annex 13, EU guidelines ect). c) International scientific trial specific standards d) Our internal requirements (e.g. internal requirements to conduct trials with high risks (phase I trials), which go further than what is specified e.g. in the law) (…)
Supranational and national patient organizations		
National Breast Cancer Coalition and Nancy Roach, Colorectal Cancer Coalition (USA)	USA	(…)what "quality" means, i.e., what truly matters, to patients themselves: • (…)"quality" and risk-based quality management requires patient-centred clinical trials that are scientifically valid and designed to robustly, efficiently answer questions of true import to patients, rather than questions that are simply of scientific interest but ultimately would have little impact on enhancing patient care. • It requires trials that are designed to prevent risks and errors that truly matter to patient safety and the validity of the trial data. • In addition, quality means patient-centred trials that appropriately incorporate patient preferences into study design and comprise "rational" design that minimises patient burden and maximises patient benefit. And from the patients' perspectives, "quality" also is defined by certain "don'ts": • quality trials are those that do not introduce invasive and/or repeated procedures, unnecessarily numerous study visits, and unnecessary costs for patients that are not required for answering the trial's questions. • They do not introduce unnecessarily restrictive inclusion and exclusion criteria that hamper accrual and may generate data that do not accurately reflect safety and efficacy for the larger patient population. • And participation in such trials does not require unneeded delays in treatment initiation secondary to screening and trial arm assignment. Coming full circle, quality trials provide uniformity in recruiting patients; are feasible and "practical" for both patients and their providers; include patient-centred, patient-friendly informed consents that truly inform patients; continually keep trial participants informed—whether the results are positive or negative; and move our body of knowledge forward and/or change practice.

S4 Table. Quality themes appearing in proposal evaluation criteria of funding agencies

A) Overview of quality criteria used by national funding agencies

Overarching quality themes	AU	AUS	CAN	CH	DE	UK	USA	NOR	Horizon 2020
Quality of researcher, applicant, and/or team	x	x	x	x	x	x	x	x	
Impact and/or significance of research project	x		x	x		x	x	x	x
Quality of scientific approach and/or methods	x		x	x		x	x	x	
Research infrastructure & financial capacity at applying site			x		x	x	x	x	x
Consideration of diversity & equal opportunities of applicants		x			x			x	
Ethical considerations of research project	x					x		x	
Originality/innovation of research project			x	x			x		
Feasibility of research project	x			x					
Requested resources or funding adequate for proposed project					x	x			
Quality of international / national cooperation arrangements					x			x	
Dissemination & communication of results								x	

Funding agencies by country/region. Australia: Australian Government; National Health and Medical Research Council; Austria: Wissenschaftsfonds FWF (Fonds zur Förderung der wissenschaftlichen Forschung); Canada: Canadian Institutes of Health Research; Germany: Deutsche Forschungsgemeinschaft ; United Kingdom: Medical Research Council ; United States of America: National Institutes of Health ; Norway: The Research Council of Norway; Switzerland: Swiss National Science Foundation; Europe: Horizon2020

B) Examples of criteria used by funding agencies to evaluate research proposals, by quality theme

Overarching quality themes	Examples (quotes)
Quality of researcher, applicant, and/or team	*"Qualifications of the applicant(s), including training, experience and independence (relative to career stage).* - *Experience of the applicant(s) in the proposed area of research and with the proposed methodology.* - *Expertise of the applicant(s), as demonstrated by scientific productivity over the past five years (publications, books, grants held, etc.). Productivity should be considered in the context of the norms for the research area, applicant experience and total research funding of the applicant.* - *Ability to successfully and appropriately disseminate research findings, as demonstrated by knowledge translation activities (publications, conference presentations, briefings, media engagements, etc.).* - *Appropriateness of the team of applicants (if more than one applicant) to carry out the proposed research, in terms of complementarity of expertise and synergistic potential." (**Canadian Institutes of Health Research)*** *"Applicants must have the professional competencies and qualifications required to complete the proposed action or work programme: it may be assessed on the basis of specific qualifications, professional experience and references in the field concerned" **(Horizon2020)***
Impact and/or significance of research project	*"Assess the potential economic and social impact of the proposed research including:* - *Identification of realistic improvements to human or population health* - *Contribution to relieving disease/disability burden and/or improving quality of life* - *Identification of potential impacts of research and plans to deliver these" **(Medical Research Council, UK)*** *"Does the project address an important problem or a critical barrier to progress in the field? If the aims of the project are achieved, how will scientific knowledge, technical capability, and/or clinical practice be improved? How will successful completion of the aims change the concepts, methods, technologies, treatments, services, or preventative interventions that drive this field? " **(National Institutes of Health, USA)***
Quality of scientific approach and/or methods	*"This criterion gives an indication of the essential, fundamental aspects of the research project. The scientific merit of a project will be assessed in relation to the following points:* - *Originality in the form of scientific innovation and/or the development of new knowledge.* - *Whether the research questions, hypotheses and objectives have been clearly and adequately specified.* - *The strength of the theoretical approach, operationalisation and use of scientific methods.* - *Documented knowledge about the research front.* - *The degree to which the scientific basis of the project is realistic.* - *The scientific scope in terms of a multi- and interdisciplinary approach, when relevant." **(The Research Council of Norway)***
Research infrastructure & financial capacity at applying site	*"Availability and accessibility of personnel, facilities and infrastructure required to conduct the research.* - *Suitability of the environment to conduct the proposed research.* - *Suitability of the environment (milieu, project and mentors) for the training of personnel (if applicable)."*

	(Canadian Institutes of Health Research) *"Will the scientific environment in which the work will be done contribute to the probability of success? Are the institutional support, equipment and other physical resources available to the investigators adequate for the project proposed? Will the project benefit from unique features of the scientific environment, subject populations, or collaborative arrangements?"* ***(National Institutes of Health, USA)***
Consideration of diversity & equal opportunities of applicants	*"Proposal reviews should not disadvantage applicants due to extra-scientific reasons, such as age, gender or disability. Consider the applicant's scientific career development rather than his/her age. You may compensate for certain extra-scientific disadvantages; unavoidable delays in the applicant's scientific career (for example childcare responsibilities causing longer periods of qualification, gaps in publications, or less time spent abroad) should be taken into consideration."* ***(Deutsche Forschungsgemeinschaft, Germany)***
Ethical considerations of research project	*"Does the project give rise to any ethical issues?* ***(Wissenschaftsfonds FWF, Austria)***
Originality/innovation of research project	*"Originality of the research question"* ***(Swiss National Science Foundation)***
Feasibility of research project	*"(…) and feasibility of the proposal in terms of strengths and weaknesses"* **(*Wissenschaftsfonds FWF, Austria)***
Requested resources or funding adequate for proposed project	*"Justification of the proposed staff needs by the work programme* *(…) Necessity and utilisation of the proposed instruments"* ***(Deutsche Forschungsgemeinschaft, Germany)***
Quality of international / national cooperation arrangements	*For special programs: "Quality and add-on value of cooperation arrangements"* ***(Deutsche Forschungsgemeinschaft, Germany)*** *Additional criteria: "Quality of national and international cooperation"* ***(The Research Council of Norway)***
Dissemination & communication of results	*"This criterion gives an indication of the quality of the dissemination and communication plans for the project. Dissemination and communication of results will be assessed in relation to the following points:* *• Plans for scholarly publication, dissemination and other communication activities.* *• Plans for popular science dissemination and communication activities vis-à-vis the general public as well as users of the project results, including planned use of channels and measures.* *• Plans for ensuring that important users (in industry, community life and public administration) are incorporated into/take part in dissemination activities for the project.* *When assessing dissemination and communication plans, importance should be attached to the level of detail provided and how realistic the plans are****." (The Research Council of Norway)***

Manuscript II: Towards increasing value and reducing waste in academic clinical research: Consensus on a comprehensive framework of clinical research quality

Belinda von Niederhäusern[1], MSc; Gordon H. Guyatt[3], Prof.; Matthias Briel[2,3]*, PD; and Christiane Pauli-Magnus, Prof.[1]*

[1]Clinical Trial Unit, Department of Clinical Research, University Hospital Basel, Basel, Switzerland (Emails: belinda.vonniederhausern@usb.ch; Christiane.pauli-magnus@usb.ch)

[2]Basel Institute for Clinical Epidemiology and Biostatistics, Department of Clinical Research, University Hospital Basel, Basel, Switzerland (Emails: Matthias.briel@usb.ch)

[3]Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, ON, Canada (Emails: guyatt@mcmaster.ca)

*shared senior authorship

Status: Final draft, submission planned End of August 2017

Research in context

Evidence before this study

In 2014, *The Lancet* published a Series ("Increasing value: reducing waste") providing a voluminous body of evidence for sources of waste in biomedical research, along with 17 recommendations on how to increase value, covering various stakeholders including funders, regulators, journals, academic institutions, and researchers. In 2016, a follow-up article emphasized little recognition of these recommendations by academic institutions. In a recent systematic survey of the literature and clinical research stakeholders, we showed that a commonly agreed definition or concept of "high quality research" remains to be developed.

Added value of this study

We suggest the first comprehensive framework that allows for assessment of the quality of academic clinical research. It is based on the consensus of a range of stakeholder groups across geographic regions, with a focus on its application in the academic setting. For the first time, varying stakeholder perspectives on quality are distilled in six quality dimensions that apply to all study designs involving patients. The framework provides guidance on how to assess quality at any point in a clinical study, i.e. from conceptualization of the research question to dissemination of study results. Through its development, major stakeholder groups have agreed on a common, holistic approach to quality and how to ultimately improve it in the future.

Implications of the available evidence

This framework may facilitate efforts to reduce waste and increase value, primarily as a common structure for the assessment of quality of clinical research at academic institutions. In a first step, its development has triggered all national stakeholders in Switzerland to convene and discuss the way forward to improve value in clinical research. It is in this setting that the framework currently undergoes refinement until it is validated for practice. Other interested stakeholder groups may validate the framework for their settings and share their approaches, thereby contributing to a growing momentum of increasing value and reducing waste.

Abstract

Background: A 2014 Lancet series suggested that a high proportion of investment in biomedical research is wasted, and this waste is avoidable. Academic institutions have, thus far, shown little attention to recommendations for increasing value and reducing waste. This study aimed to develop a conceptual framework guiding the comprehensive assessment of clinical research quality at academic institutions that could facilitate adoption of waste-reducing strategies.

Methods: Based on a systematic survey of quality definitions, concepts, and criteria in the medical literature and on stakeholder websites from 12 countries, we systematically developed a comprehensive framework for clinical research quality. We conducted four rounds of an adapted online Delphi process among eight stakeholder groups from 16 countries that ultimately achieved consensus on structure and content.

Findings: All 52 final Delphi respondents agreed on an overall framework structure. The framework spans five study stages (concept, planning and feasibility, conduct, analysis and interpretation, reporting and dissemination) and includes the following dimensions: (1) protection of participants' safety and rights, (2) relevance and patient centeredness and involvement, (3) minimization of bias / internal validity, (4) precision, (5) transparency/public access to data, and (6) generalizability of study results. These dimensions are interacting with two promotors, education and infrastructure, that include a set of factors that may enhance all listed quality dimensions. Each quality dimension contains main questions and explanatory items that guide quality assessment at each individual research stage from conceptualization of the research question through dissemination of study results. Between 96.2% (50/52) and 100% of Delphi participants agreed on content and wording of these 76 main questions, depending on the research stage.

Interpretation: We propose the first consensus-based framework guiding the assessment of quality of clinical research for academic research. Operationalization of this guidance will support the reduction of waste, from posing the right research question to the transparent publication of results.

Introduction

Clinical research should generate reliable evidence to best inform decision-making in clinical practice and health policy, considering benefits, harms and cost.[1] Evidence on sources and extent of waste in research has, however, highlighted imbalanced research question selection, poor study design and execution, as well as non-publication and selective reporting.[2-6] Clinical research stakeholders have expressed concerns that the current model for conducting studies is unaffordable, unsustainable and, for the generation of new knowledge, seriously flawed.[1, 7-13] Low quality clinical research may not only result in invalid data or distorted outcomes [14], but is also unethical and compromises patients' safety and rights.

In 2014, *The Lancet* published a Series ("Increasing value, reducing waste") [8-13] providing a voluminous body of evidence for sources of waste in biomedical research that also apply to clinical research. Along with a detailed analysis of potential sources of waste, the authors made 17 recommendations on how to increase value, covering various stakeholders including funders, regulators, journals, academic institutions, and researchers. A follow-up article in 2016 offered an overview of the initial stimulus of the series across stakeholder communities.[2] Although the authors noted innovation and momentum for corrective actions by some stakeholder groups, they specifically emphasized little recognition of the series by academic institutions.

As a major driving force of patient-oriented clinical research, academia would be ideally placed to lead the movement to reduce waste. Academia not only receives large proportions of public funding [15, 16] but is also producing the majority of scientific publications.[17] Despite significant investments in infrastructure, training, and methodological support [18, 19], the issues raised by the Lancet authors persist. These include financial conflicts of interest [20] and scientific misconduct [21], limited dissemination of clinical trial results [22], and a higher risk for discontinuation in public versus industry-funded trials.[23] [24] Potential reasons for slow progress and uptake of the Lancet recommendations include a lack of common academic policies across a complex ecosystem of stakeholders and their agendas.[2] Medical specialties and expert groups, ethics committees, regulatory bodies, funding agencies, industrial partners, and patients all have a say in academic research. Their perceptions on what constitutes "good clinical research", however, are vague and, to the extent they are articulated at all, vary.[25] Lack of a prior common understanding of the pillars that frame good clinical research, and practical guidance on how to improve the current situation, seriously inhibits efforts within the academic system to increase value.

The aim of this study was therefore to formulate an academic response to *the Lancet* series by (i) achieving consensus across a wide range of stakeholder groups on a comprehensive framework for the quality of clinical research, and (ii) developing guidance on how to operationalize the framework. We focus on clinical research conducted with patients and take the perspective of high-income countries.

As a model, we highlight examples of the first successful applications of the framework in the Swiss academic setting. Switzerland has a long-standing tradition of initiatives that aim to improve the quality of academic clinical research. [26] Most prominently, Clinical Trial Units (CTUs) at all University hospitals partner with national funders, regulatory bodies, and policy makers, but also international initiatives such as the European Patient's Academy on Treatment Innovation (EUPATI) or the European Clinical Research Infrastructures Network (ECRIN)[27] to continuously improve academic research.[28] It is in this context that we used the framework to establish consensus on the way forward to increase value of clinical research.

Methods

We developed a framework guided by the following principles

i) integrating available empirical evidence on quality through a systematic survey [25] informing a first matrix of quality dimensions,
ii) including the views of a broad range of stakeholder representatives[29-32] through four iterative rounds of a modified online Delphi process following current guidelines[33-35],
iii) Addressing operationalization of the framework through detailed feedback of stakeholders from the Swiss academic setting.

The scope of the framework covers different types and phases of clinical research. In the context of this work, we define clinical research as research conducted with patients to answer therapeutic, preventive, diagnostic, or prognostic questions, investigations of the mechanisms of human disease, or the development of new technologies.

Appendix A presents a detailed description of the framework development and the consensus process. In short, we consolidated the definitions, criteria, and themes as well as the derived tools and checklist identified through our systematic survey [25] into a comprehensive framework matrix applying the framework method according to Gale.[36] We

first coded quality definitions, criteria, or themes into quality items (i.e. single aspects of quality) and grouped them thematically and according to stages of research. We conducted iterative consultation addressing the comprehensiveness and presentation of the framework until we reached internal consensus.

Delphi process

Our team, with help from affiliated collaborators, and by word of mouth among the related networks (e.g. European Patient Academy on Therapeutic Innovation (EUPATI) for patient representatives) identified potential stakeholder representatives from 16 countries and seven groups (patient organizations and representatives, academic researchers/initiatives, medical faculties and clinical trial units, governmental bodies, regulatory agencies, ethics committees, the pharmaceutical industry, and funding agencies). We recruited participants on the basis of awareness of quality issues related to clinical research and ability to provide feedback within a specified time window. After invitation of representatives through the survey software SurveyMonkey© (www.surveymonkey.net), we conducted two Delphi-rounds aimed at (i) identifying any additional quality item that we had not yet considered, and (ii) establishing broad consensus across stakeholders on the overall framework structure. Consensus was pre-defined as an agreement of 80% or higher. After each round, we shared with respondents a summary of the adaptations made based on their suggestions in the previous round and asked for their agreement or further improvements and suggestions on structure and content.

Seeking consensus on how to operationalize the framework structure in the Swiss academic setting, for Delphi rounds three and four, we invited additional stakeholder representatives from Switzerland, particularly academics. This included consensus on framework structure, content, and wording of main quality questions and corresponding examples. In round four, we additionally provided respondents with all anonymized comments, a response by the authors to each comment, and the overall agreement score on framework structure and main quality questions.

Data Analysis

Qualitative analysis was done for open-ended questions. We descriptively analyzed comments and suggestions for removal, addition, or adaptation of quality dimensions, and individual quality items and identified key themes based on repetition of concept words. Through discussions amongst the authors, the framework structure and content was adapted iteratively and fed back to the survey participants for discussion. Agreement scores were

calculated in percentages by dividing the number of participants agreeing by the total of participants who provided an answer to the respective question.

Role of the funding source

This research did not receive grants from funding agencies in the public, commercial, or not-for-profit sectors. It was fully funded by an in-house grant of the Department of Clinical Research, which is co-chaired by Prof. Christiane Pauli-Magnus (CPM). CPM was involved in study design, collection, analysis and interpretation of the data, in writing the report, and in the decision to submit the paper for publication. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

Description of framework structure

We invited 109 stakeholder representatives from 16 countries to participate in our Delphi process (**Appendix A, Tables 1 and 2** present response rates). Total response rates ranged from 53.0% (58/109) in round one to 96.3% (52/54) in round four, and from 12.5% (3/24) for governmental representatives to 83.3% (10/12) of pharmaceutical industry representatives in the first round. In all four rounds, academic representatives were overrepresented to allow wide participation. After an agreement of 53.1% (26/49) of respondents in round one, 97.1% (33/34) in round two and 87.0% (47/54) in round three, 100% (52/52) of survey respondents agreed on a final framework structure (i.e. all building blocks, wording, and order) in round four. The final agreement on content, i.e. the 76 main questions and their wording ranged from 96.2% (50/52, sustainability and education) and 98.1% (51/52, planning and feasibility) to 100% in all other stages. **Appendix B** presents all versions of the framework with track changes and the author's reply to comments**.** Three main building blocks provide the final framework is structured into (Figure 1):

a) Six quality dimensions, with a dimension being defined as an overarching concept of quality containing multiple individual quality questions (Box 1): (1) Protection of patient safety and rights; (2) Relevance of study question and patient centeredness and involvement; (3) Minimization of bias, i.e. internal validity; (4) Precision; (5) Transparency and access to data; and (6) Generalizability, i.e. external validity of study,

b) five successive study stages (concept, planning and feasibility, conduct, analysis and interpretation, and reporting and dissemination) to which the dimensions apply, and
c) two quality promoters (infrastructure, sustainability and education), with a promotor being defined as set of factors that may enhance all listed quality dimensions at a research institution.

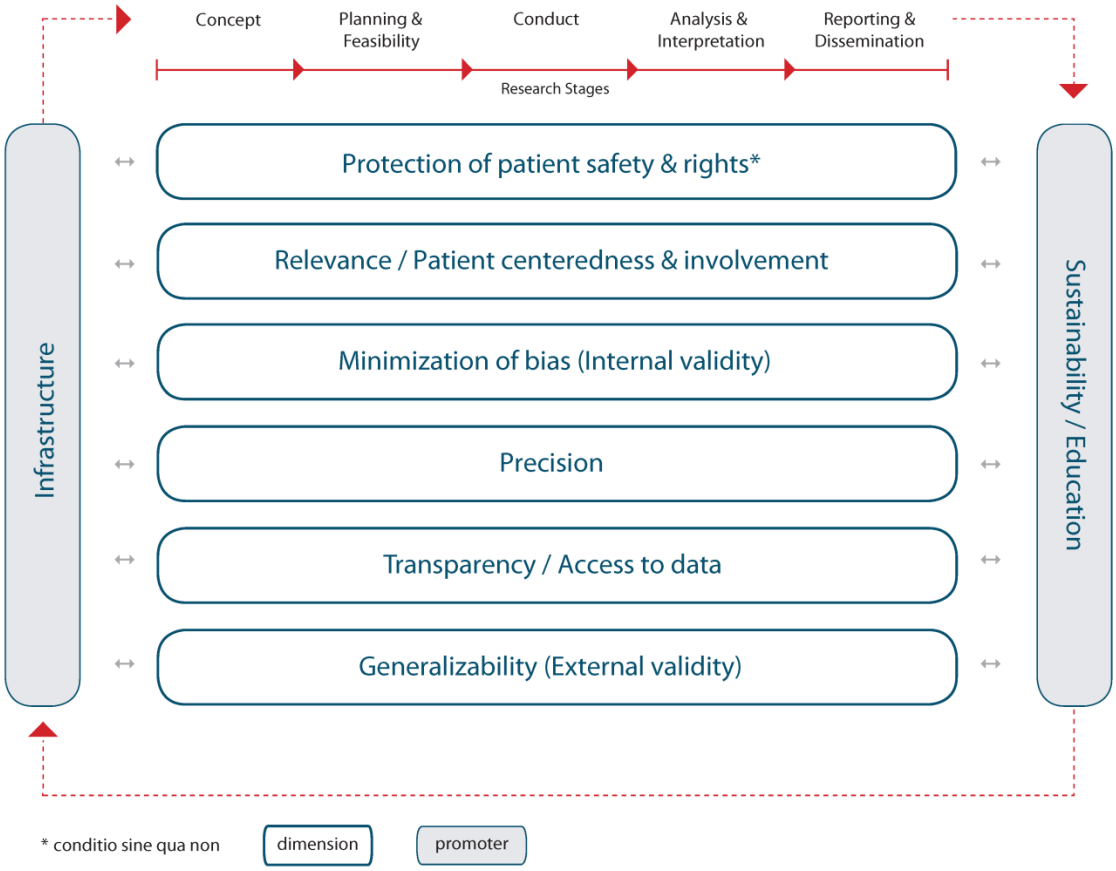


Figure 1. Structure of the framework for quality of clinical research. Box 1 provides a brief description of each dimension and promoter.

Box 1. Description of quality dimensions and promoters

Quality dimensions

- The first quality dimension, *protection of patients' safety and rights*, represents the cornerstone of clinical research. It assures that participants' safety, rights, and well-being are respected and protected at all times.
- The second quality dimension, *research relevance and patient centeredness and involvement* reflects the extent to which the research question is scientifically and societally beneficial (i.e. leads to improved decision-making in health care) and involves patient values and preferences at all stages.
- The third quality dimension, *minimization of bias – or internal validity*, reflects the extent to which systematic error (bias) is minimized, i.e. through selecting an appropriate study design and pre-specifying analyses.
- The fourth quality dimension, *precision - or statistical validity*, reflects the extent to which random error is minimized (i.e. sufficiently narrow confidence intervals are achieved to confirm or reject clinical hypothesis), and to what level precision is reported and described in order for readers to be able to judge it.
- The fifth quality dimension, *transparency and access to data*, reflects the extent to which study planning, conduct, data collection and presentation of results are transparent to and accessible for the scientific community and the public. It includes the registration of the study in a publicly accessible database, publication of the full study protocol, publication of the study results - independent of their effect size or direction-, and explicitly, encouraging access to the full patient-level data set (data sharing).
- The sixth quality dimension, *generalizability – or external validity*-, reflects the extent to which study results are applicable and generalizable to the wider patient population in real life circumstances.

Quality promoters

- The first quality promoter consists of an established research *infrastructure* with well-trained personnel and functional facilities on-site.
- The second quality promoter supports *sustainability* of a developed infrastructure through effective involvement and hands-on training of young and senior investigators as well as competent mentoring and early career development, and continuous *education* of study personnel in order to secure a productive clinical research environment in the long term.

For a detailed description of the dimensions and promoters and empirical evidence supporting their importance, please see **Appendix C**.

Operationalization of the framework

In contrast to the existing quality assessment tools or checklists [25], we aimed to develop a framework in which all quality dimensions are applicable to each step in the conduct of clinical studies. During Delphi round one and two, participants commented on the characterization of study stages and ultimately achieved consensus. The first study stage, *conceptualization*, starts with a clinical knowledge gap and ends with a clearly defined research question and appropriate study design. During the second study stage, *planning and feasibility*, the investigators develop a protocol based on their research question and assess the feasibility of their undertaking. This stage ends with approval of the protocol by regulatory bodies (if such approval is necessary). The third stage, *study conduct,* starts with the first patient recruited and ends with last patient last visit. During this stage, the investigators conduct the study according to the approved protocol. In stage four, *analysis and interpretation*, the investigators process and interpret the study data generated during conduct of the study. The last study stage, *reporting and dissemination*, covers all activities after analysis of the data, i.e. publication, dissemination, and archiving of study results.

Applying the framework in the academic clinical research context (e.g. the longitudinal quality assessment of a particular study, or a study's potential for funding), requires guidance on its operationalization. To this end, we developed "main quality questions" on which for each stage the Delphi participants commented, and on which – following revision - they finally agreed (see **Appendix B for all versions of the framework and Appendix C, Table 1 for agreement scores**). These main quality questions illustrate what quality aspect should be addressed and answered in a specific study stage. As an example, during concept phase, the dimension "Relevance / patient centeredness and involvement" can be further specified by asking the main question "Is significant add-on value to already existing evidence given, taking into consideration burden of disease and anticipated benefit of treatment?" (**Table 1**). Although the questions are tailored to the academic setting in a high-income country, all stakeholder groups agreed on these questions in Delphi round four. The main questions are explicitly designed to be broadly applicable to different study designs. More specific guidance on how to operationalize the framework for a particular study type or setting is provided by *descriptive examples* complementing the main questions in **Appendix D.**

The development of the framework has already triggered several Delphi participants and members of the Swiss Clinical Trial Organization (SCTO) to organize a first Swiss symposium on "Adding value in clinical research: what's been achieved and how do we manage new challenges? in June 2017 (https://www.scto.ch/de/event-

calendar/symposium/symposium-2017.html). Along with two authors of the Lancet series, the current situation in Switzerland was discussed and the quality framework content presented.

Table 1. Content of the framework for the quality of clinical research including main questions, by research stage. Appendix D presents the full framework including quality promoters and examples**.**

Study Stage I: Concept **Milestone: Research question including study type defined and viable**	
Dimension	**Main question**
Protection of patient safety and rights	Can the research question be answered in the given setting?
	Does study consider equity appropriately?
	Is the research design adequate for the stage of an investigated technology to ensure patient safety?
	Do the (assumed) short and long term benefits of the study outweigh potential risks associated with the study (consistent with clinical equipoise)?
Relevance / **Patient centeredness and involvement**	Is significant add-on value to already existing evidence given, taking into consideration burden of disease and anticipated benefit of treatment?
	Are patient representatives/ advocates and their needs and values adequately involved in the development of the research question?
	Are outcome measures patient-relevant?
Minimization of bias (internal validity)	Is the selected study type/design appropriate to minimize bias?
	Are potential sources of bias anticipated, evaluating the magnitude and the likely direction?
	Are outcome measures well-defined, pre-specified, valid, reliable and measured at appropriate times?
Precision	Has estimate of the required sample size been made (for feasibility purposes, see "Protection of patient safety & rights")?
Transparency / Access to data	Is the research question clearly specified (e.g. in a synopsis)?
Generalizability (external validity)	Are planned study participants representative of patients who would use the drug/intervention/diagnostic test in a real-life setting?
Study Stage II: Planning and Feasibility **Milestone: Protocol developed and approved by regulatory bodies**	
Dimension	**Main question**
Protection of patient safety and rights	Do the potential short and long term benefits of the study outweigh study burden (due to study visits, intervention, procedures etc.)?
	Are patients' safety and rights protected through the study's adherence to applicable national and international regulations and laws?
	Has feasibility been checked thoughtfully based on existing evidence?
	Is collection, documentation, and reporting of Adverse Events / Serious Adverse Events / Suspected Unexpected Serious Adverse Reaction according to the applicable regulations planned and specified in the protocol?
	Are mechanisms established which allow early study termination when required and prevent early study termination for inadequate reasons?
Relevance / **Patient centeredness and**	Is knowledge transfer/use (e.g. plans for inclusion of results in clinical guidelines) planned?

involvement	
Minimization of bias (internal validity)	Is statistical analysis pre-specified (using outcomes as defined in concept stage)?
	Is study monitoring (adapted to risk of study, if applicable) planned and documented in a monitoring plan??
	Is data management planned and documented in a data management plan?
	Is minimization of bias planned for according to the research question and study design?
Precision	Are expected treatment effects and event rates in intervention and control groups realistic and estimated based on empirical evidence?
	Are recruitment procedures and recruitment monitoring planned to ensure sufficient sample size?
Transparency / Access to data	Is the protocol in accordance with established standards (e.g. SPIRIT [37] or other applicable guidelines depending on study design)?
	Is there a dissemination plan to share study information including the protocol, summary results, and participant level data?
Generalizability (external validity)	Are study procedures/observations in line with routine practice in the given setting?
Study Stage III: Conduct **Milestone: Last patient last visit**	
Dimension	**Main question**
Protection of patient safety and rights	Is respect for and consideration of patient rights, well-being and dignity guaranteed throughout conduct of study?
	Is patient safety guaranteed throughout conduct of study?
	Is study conducted according to protocol?
	Is compliance of participants and study staff with protocol monitored?
	Are patients' safety and rights protected through the study's adherence to applicable national and international regulations and laws?
Relevance / Patient centeredness and involvement	Are there any measures in place to assure study participants' involvement, cooperation, and feedback throughout conduct of study (e.g. incentives, phone calls, etc.)?
Minimization of bias (internal validity)	Are data systematically collected as pre-specified in protocol?
	Is monitoring conducted according to the pre-specified monitoring plan?
Precision	Is enrollment of study participants monitored?
	Is variability of study procedures and measurement error minimized, e.g. by utilizing centralized monitoring strategies?
Transparency / Access to data	Is study conduct transparent to all involved parties?
Generalizability (external validity)	Are numbers of participants through different stages of a study documented (patient flow) including reasons for leaving the study before its end (if voluntarily provided by patient)?

Study Stage IV: Analysis and Interpretation **Milestone: Study data analyzed and interpreted**	
Dimension	**Main question**
Protection of patient safety & rights	Does data sharing adhere to appropriate data protection policies?
Relevance / **Patient centeredness**	Are data analyzed so that the use of results by different stakeholders is maximized?
Minimization of bias (internal validity)	Is the data analyzed as pre-specified in the protocol/statistical analysis plan?
	Are key confounding variables adjusted for in the analysis (e.g. multivariable analysis), if applicable?
	Does the analysis follow an adequate strategy to deal with participants in whom treatment or follow-up was not in accordance with study protocol?
	Are results interpreted with least possible "spin"? (e.g. without intentionally implying greater or lesser effects than actually shown by the data)?
Precision	Is the uncertainty of results considered in the analysis?
Transparency / Access to data	Is the analysis code clearly documented and the analysis process reproducible?
	Are deviations from the statistical analysis plan or protocol adequately documented and reported?
Generalizability (external validity)	Does the interpretation put the results adequately into context of clinical practice/public health?
Study Stage V: Reporting and Dissemination **Milestone: Study archived and published**	
Dimension	**Main question**
Protection of patient safety and rights	Is study completion/termination communicated to appropriate parties and documented in registries?
	Are study participants informed about outcome/main findings of the study in plain language (including treatment allocation of participant, if applicable)?
	Do study participants get access to products/interventions after study, if applicable?
Relevance / **Patient centeredness and involvement**	Do authors critically reflect on research findings (results as well as challenges or mistakes during study conduct) and the implications for future research?
	Is the study easily available to decision/policy/guideline makers?
	Are study patients/patient representatives involved in the reporting of the study?
Minimization of bias (internal validity)	Are all outcomes and important study characteristics reported as pre-specified in the protocol (outcome reporting bias prevented)?
Precision	Are absolute and relative treatment effects reported accompanied by confidence intervals?
	Is the analysis set of participants clearly specified?
Transparency / Access to data	Is dissemination of data and study results maximized?
	Are reporting guidelines followed to facilitate critical appraisal and reproducibility?
	Are selective reporting, "spin", plagiarism and self-plagiarism avoided and conflicts of interest declared?

	Is knowledge transfer & exchange fostered?
	Are study records and data sets kept and archived for the legally required period of time?
Generalizability (external validity)	Is potential impact on clinical practice / public health outlined in publicly accessible research reports (e.g. journal publication)?
	Are characteristics of included participants clearly reported
	Are results of pre-specified subgroup analyses, if applicable, reported in order to assess the importance of key participant characteristics (e.g. disease severity, age or gender)?

Discussion

This study represents the first effort to involve a range of stakeholder groups across geographic regions to formulate an academic response to *the Lancet* series on increasing value and reducing waste in clinical research. We suggest a comprehensive, consensus-based quality framework that is applicable to all study types and spans the entire lifecycle of a clinical study, i.e. from conceptualization of the research question to dissemination of study results. We designed the framework to increase value through operationalization by academic institutions that, thus far, have shown only little recognition of the Lancet series.

The limited resources in the academic setting urge to emphasize that relevant research should build on what is already known, i.e. be preceded by systematic reviews.[38, 39] We particularly put emphasis on the conceptualization and planning and feasibility stage at which the protocol including research question and study design is developed, and its overall feasibility is assessed. Research questions should lead to clinically relevant information gain without influence by special interest groups (e.g. industry sponsors). [40] [9, 41] Further, feasibility assessments of the planned study are crucial in order to avoid waste in financial and human resources – and to justify exposing participants to burdens or risks. A lack of feasibility assessment often results in low recruitment rates leading to the many clinical trials that are terminated prematurely [7, 23, 42-47] that cannot generate valid scientific knowledge and are thus unethical.[48, 49] We highlight that pragmatic study designs or collection of real-world data, for example, may capture real-life circumstances in such resource-restrained settings and allow for greater applicability and consideration of external validity. [50-53]

At all stages of the framework, we provide guidance on how to involve patients from conceptualization of the research question through conduct and eventually, publication and dissemination of lay language summaries. Academic clinical research should be patient-oriented and avoid dominant commercial interest. Investigators should not only align their

research with patient priorities and the utilities patients assign to different problems and outcomes, but also ensure acceptability of their interventions.[1, 12, 54] Initiatives such as the Patient-Centered Outcomes Research Institute (PCORI) [55], the James Lind Alliance (www.lindalliance.org), and the INVOLVE Initiative in the UK (www.invo.org.uk) provide guidance and suggest patient engagement at all stages of a clinical study.

The framework and its operationalization thus potentially provide a supporting structure for the Reducing Waste, Reward Diligence (REWARD) Campaign (http://researchwaste.net/reward-statement/) to increase value and reduce waste. At Swiss national level, the framework has triggered all stakeholders to convene in a first symposium on how to increase value of academic clinical research.

Strengths and weaknesses

The strength of this study is that it is unique in achieving a consensus among very diverse stakeholder groups on a prominent, but complex and ill-defined concept. Our response and agreement rates during the two consensus finding rounds were very high. Moreover, we had prepared for this work with an extensive systematic search of the existing definitions, approaches, and measurements of clinical research quality across cultures and stakeholder groups that revealed a lack of a common concept. We therefore developed this framework based on empirical evidence and, before circulating it for consensus, considered the interests of a wide variety of stakeholders. The resulting guidance is the first of its kind to support academic institutions, researchers, and other stakeholders in the holistic assessment of study quality with the overall aim of increasing value and reducing waste in clinical research.

Our study has several limitations. In the Delphi process, participation varied across stakeholder groups and therefore, some stakeholder opinions may be under- or overrepresented. For example, patients are underrepresented because it was rather difficult for them to comment on later, more complex versions of the framework. We made every effort to contact and motivate patient representatives and to explain the framework content in plain language. In contrast, academics were intentionally overrepresented. We limited stakeholders to those who were willing to reply in English, German, French, or Italian.

We are aware that the framework with its 76 main questions covers a substantial number of topics and might be considered too comprehensive for application in practice. However, this quality framework emphasizes the need for a holistic approach to quality, rather than restricting the focus to individual dimensions. The items are not meant to be prescriptive, and

we do not define a minimum set of criteria that an investigation should meet. Although all criteria are important and, independent of the setting, any stakeholder group should address every dimension, different weights may be applied for different clinical research designs. Similarly, certain criteria or "quality promoters" may be difficult to attain in resource-restrained settings. Our collaboration with researchers from the Swiss Public and Tropical Health Institute aims to adapt the framework for lower income settings.

Next steps in implementation

The rapid uptake of the framework in Switzerland has been driven by the broad consensus across Swiss representatives of all major clinical research stakeholder groups including not only academia, but also ethics committees, regulatory and governmental bodies, funding agencies, patients, and the pharmaceutical industry. After the initial get-together at the national symposium in June, we plan a one-day strategy workshop engaging a diverse group of Swiss stakeholder representatives and policy makers at the end of this year. Together with the Swiss Clinical Trial Organization, which currently incorporates the framework content in its operational and strategic quality policies, we will discuss main objectives including how to set a quality research agenda for academic clinical research in Switzerland, how to implement measures to improve quality nationally, how to monitor progress and how to generate empirical evidence on the impact of these measures in the long run. Some of the participants will be the leaders of the Swiss initiative "We Scientists Shape Science" (https://naturwissenschaften.ch/wescientists) that held a first congress in January with over 200 scientists debating on how to improve science in Switzerland and globally.

At local level, multiple initiatives are currently ongoing to test the framework's real-world potential in Switzerland (**Box 2).** All initiatives aim at evaluating the framework's content validity through a continuous process of evaluation, refinement, and development of future versions that we will report on.

Due to its broad acceptance across international stakeholders in round one and two and its theoretical underpinning, the framework may also be utilized for setting the agenda for other stakeholder groups or geographies. **Box 1** and **2** in **Appendix E** describe a variety of scenarios at different organizational and institutional levels in which the framework may be applied.

Box 2. Increasing value of research at local level – four examples from a Swiss academic hospital

(1) Setting the research agenda

The Department of Clinical Research at the University Hospital Basel uses the framework for setting its research agenda. It systematically applies the main questions to identify "excellence potential" across research projects that apply for scarce Departmental grant funding for special methodological support.

(2) Improving methods support

The methodological support units have started to use the framework as quality guidance in their clinical/epidemiological consulting activities. Depending on the research stage of a particular study, the framework is used to guide researchers in the conceptualization, planning, conduct, analysis or dissemination of their study.

(3) Researching research

The practical challenges faced by investigators identified during consulting services serve as the basis for the methodological research units to initiate studies to investigate innovative solutions. These studies aim to generate empirical evidence on single aspects of the framework and whether they improve value or not.

(4) Monitoring impact

The Clinical Trial Unit pilots the use of the framework criteria to longitudinally asses its impact on research quality for the Swiss Clinical Trial Organisation. On site, a cardiology department has applied the framework as a tool to assess the quality of two ongoing large studies, one retrospective cohort and one RCT. The framework has also proven useful as teaching material for early career investigators in the group.

Conclusion

The proposed framework is based on the acceptance of a diversity of international and Swiss clinical research stakeholder groups with the aim to establish a common, well-differentiated academic answer to the discussion initiated by the Lancet series. The framework will, as a structure for operationalizing the assessment of quality of clinical research at academic institutions, facilitate implementation of waste reduction and value increasing initiatives. To take the field forward, we encourage the research community and interested stakeholder groups to apply the framework, generate evidence on content validity, and to transparently and openly share their approaches.

Acknowledgements

We would like to thank all experts who participated in the Delphi process and devoted their time to provide invaluable feedback.

References

[1] Ioannidis JP. Why Most Clinical Research Is Not Useful. PLoS Med. 2016;13:e1002049.
[2] Moher D, Glasziou P, Chalmers I, Nasser M, Bossuyt PM, Korevaar DA, et al. Increasing value and reducing waste in biomedical research: who's listening? Lancet. 2016;387:1573-86.
[3] Munafò MR, Nosek BA, Bishop DVM, Button KS, Chambers CD, Percie du Sert N, et al. A manifesto for reproducible science. Nature Human Behaviour. 2017;1:0021.
[4] Button KS, Ioannidis JP, Mokrysz C, Nosek BA, Flint J, Robinson ES, et al. Power failure: why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci. 2013;14:365-76.
[5] John LK, Loewenstein G, Prelec D. Measuring the prevalence of questionable research practices with incentives for truth telling. Psychol Sci. 2012;23:524-32.
[6] Kerr NL. HARKing: hypothesizing after the results are known. Pers Soc Psychol Rev. 1998;2:196-217.
[7] Kasenda B, von Elm E, You J, Blumle A, Tomonaga Y, Saccilotto R, et al. Prevalence, characteristics, and publication of discontinued randomized trials. JAMA. 2014;311:1045-51.
[8] Al-Shahi Salman R, Beller E, Kagan J, Hemminki E, Phillips RS, Savulescu J, et al. Increasing value and reducing waste in biomedical research regulation and management. Lancet. 2014;383:176-85.
[9] Chalmers I, Bracken MB, Djulbegovic B, Garattini S, Grant J, Gulmezoglu AM, et al. How to increase value and reduce waste when research priorities are set. Lancet. 2014;383:156-65.
[10] Chan AW, Song F, Vickers A, Jefferson T, Dickersin K, Gotzsche PC, et al. Increasing value and reducing waste: addressing inaccessible research. Lancet. 2014;383:257-66.
[11] Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, Julious S, et al. Reducing waste from incomplete or unusable reports of biomedical research. Lancet. 2014;383:267-76.
[12] Ioannidis JP, Greenland S, Hlatky MA, Khoury MJ, Macleod MR, Moher D, et al. Increasing value and reducing waste in research design, conduct, and analysis. Lancet. 2014;383:166-75.
[13] Macleod MR, Michie S, Roberts I, Dirnagl U, Chalmers I, Ioannidis JP, et al. Biomedical research: increasing value, reducing waste. Lancet. 2014;383:101-4.
[14] Juni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. BMJ. 2001;323:42-6.
[15] Dorsey ER, de Roulet J, Thompson JP, Reminick JI, Thai A, White-Stellato Z, et al. Funding of US biomedical research, 2003-2008. Jama. 2010;303:137-43.
[16] Moses H, 3rd, Matheson DH, Cairns-Smith S, George BP, Palisch C, Dorsey ER. The anatomy of medical research: US and international comparisons. Jama. 2015;313:174-89.
[17] Ioannidis JP, Boyack KW, Klavans R. Estimates of the continuously publishing core in the scientific workforce. PLoS ONE. 2014;9:e101698.
[18] Galsworthy MJ, Hristovski D, Lusa L, Ernst K, Irwin R, Charlesworth K, et al. Academic output of 9 years of EU investment into health research. The Lancet.380:971-2.
[19] Heinig SJ, Dev A, Bonham AC. The U.S. Public׳s Investment in Medical Research: An Evolving Social Contract. The American Journal of the Medical Sciences.351:69-76.
[20] Lo B. Serving two masters--conflicts of interest in academic medicine. N Engl J Med. 2010;362:669-71.
[21] Binder R, Friedli A, Fuentes-Afflick E. The New Academic Environment and Faculty Misconduct. Acad Med. 2016;91:175-9.
[22] Chen R, Desai NR, Ross JS, Zhang W, Chau KH, Wayda B, et al. Publication and reporting of clinical trial results: cross sectional analysis across academic medical centers. Bmj. 2016;352.
[23] Kasenda B, von Elm E, You J, Blumle A, Tomonaga Y, Saccilotto R, et al. Prevalence, characteristics, and publication of discontinued randomized trials. Jama. 2014;311.

[24] Meador KJ. Decline of clinical research in academic medical centers. Neurology. 2015;85:1171-6.
[25] von Niederhäusern B, Schandelmaier S, Mi Bonde M, Brunner N, Hemkens LG, Rutquist M, et al. Towards the development of a comprehensive framework: Qualitative systematic survey of definitions of clinical research quality. PLoS ONE. 2017;12:e0180635.
[26] Akert K, Fey H, Gautier E. Bericht "Medizin Schweiz", Koordination der Forschungsfinanzierung: Standortbestimmung bestehender sowie Förderung der Bildung neuer Kompetenzzentren biomedizinischer Grundlagen-und klinischer Forschung. Bern1994.
[27] Guidelines for Good Operational Practice V2.0. Swiss Clinical Trial Organisation SCTO; 2013.
[28] von Niederhausern B, Fabbro T, Pauli-Magnus C. The role of Clinical Trial Units in investigator- and industry-initiated research projects. Swiss Med Wkly. 2015;145:w14161.
[29] Campbell SM, Cantrill JA, Roberts D. Prescribing indicators for UK general practice: Delphi consultation study. Bmj. 2000;321:425-8.
[30] Murphy E, Dingwall R, Greatbatch D, Parker S, Watson P. Qualitative research methods in health technology assessment: a review of the literature. Health Technol Assess. 1998;2:iii-ix, 1-274.
[31] Murphy MK, Black NA, Lamping DL, McKee CM, Sanderson CF, Askham J, et al. Consensus development methods, and their use in clinical guideline development. Health Technol Assess. 1998;2:i-iv, 1-88.
[32] Elwyn G, O'Connor A, Stacey D, Volk R, Edwards A, Coulter A, et al. Developing a quality criteria framework for patient decision aids: online international Delphi consensus process. Bmj. 2006;333:417.
[33] Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. J Adv Nurs. 2000;32:1008-15.
[34] Sinha IP, Smyth RL, Williamson PR. Using the Delphi technique to determine which outcomes to measure in clinical trials: recommendations for the future based on a systematic review of existing studies. PLoS Med. 2011;8:e1000393.
[35] Boulkedid R, Abdoul H, Loustau M, Sibony O, Alberti C. Using and reporting the Delphi method for selecting healthcare quality indicators: a systematic review. PLoS ONE. 2011;6:e20476.
[36] Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC Med Res Methodol. 2013;13:117.
[37] Chan AW, Tetzlaff JM, Gotzsche PC, Altman DG, Mann H, Berlin JA, et al. SPIRIT 2013 explanation and elaboration: guidance for protocols of clinical trials. Bmj. 2013;346:e7586.
[38] Clarke M, Chalmers I. Discussion sections in reports of controlled trials published in general medical journals: islands in search of continents? Jama. 1998;280:280-2.
[39] Clarke M, Hopewell S, Chalmers I. Reports of clinical trials should begin and end with up-to-date systematic reviews of other relevant evidence: a status report. J R Soc Med. 2007;100:187-90.
[40] Minelli C, Baio G. Value of Information: A Tool to Improve Research Prioritization and Reduce Waste. PLoS Medicine. 2015;12:e1001882.
[41] Evangelou E, Siontis KC, Pfeiffer T, Ioannidis JP. Perceived information gain from randomized trials correlates with publication in high-impact factor journals. J Clin Epidemiol. 2012;65:1274-81.
[42] Blumle A, Schandelmaier S, Oeller P, Kasenda B, Briel M, von Elm E. Premature Discontinuation of Prospective Clinical Studies Approved by a Research Ethics Committee - A Comparison of Randomised and Non-Randomised Studies. PLoS ONE. 2016;11:e0165605.
[43] Briel M, Olu KK, von Elm E, Kasenda B, Alturki R, Agarwal A, et al. A systematic review of discontinued trials suggested that most reasons for recruitment failure were preventable. J Clin Epidemiol. 2016;80:8-15.
[44] Rosenthal R, Kasenda B, Dell-Kuster S, von Elm E, You J, Blumle A, et al. Completion and publication rates of randomized controlled trials in surgery: an empirical study. Ann Surg. 2015;262.

[45] Schandelmaier S, Tomonaga Y, Bassler D, Meerpohl JJ, von Elm E, You JJ, et al. Premature Discontinuation of Pediatric Randomized Controlled Trials: A Retrospective Cohort Study. J Pediatr. 2017;184:209-14 e1.
[46] Schandelmaier S, von Elm E, You JJ, Blumle A, Tomonaga Y, Lamontagne F, et al. Premature Discontinuation of Randomized Trials in Critical and Emergency Care: A Retrospective Cohort Study. Crit Care Med. 2016;44:130-7.
[47] Stegert M, Kasenda B, von Elm E, You JJ, Blumle A, Tomonaga Y, et al. An analysis of protocols and publications suggested that most discontinuations of clinical trials were not based on preplanned interim analyses or stopping rules. J Clin Epidemiol. 2016;69:152-60.
[48] Laurence DR. Ethics and Regulation of Clinical Research. J Med Ethics. 1988;14:44-6.
[49] Emanuel EJ, Wendler D, Grady C. What makes clinical research ethical? JAMA. 2000;283:2701-11.
[50] Tunis SR, Stryer DB, Clancy CM. Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. JAMA. 2003;290:1624-32.
[51] Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?". Lancet. 2005;365:82-93.
[52] Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. The PRECIS-2 tool: designing trials that are fit for purpose. Bmj. 2015;350:h2147.
[53] Treweek S, Zwarenstein M. Making trials matter: pragmatic and explanatory trials and the problem of applicability. Trials. 2009;10:37.
[54] Mullins CD, Vandigo J, Zheng Z, Wicks P. Patient-centeredness in the design of clinical trials. Value Health. 2014;17:471-5.
[55] Selby JV, Lipstein SH. PCORI at 3 years--progress, lessons, and plans. N Engl J Med. 2014;370:592-5.
[56] Bassler D, Briel M, Montori VM, Lane M, Glasziou P, Zhou Q, et al. Stopping randomized trials early for benefit and estimation of treatment effects: systematic review and meta-regression analysis. Jama. 2010;303:1180-7.
[57] Higgins JP, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. Bmj. 2011;343:d5928.
[58] Sterne JA, Hernan MA, Reeves BC, Savovic J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. Bmj. 2016;355:i4919.
[59] Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155:529-36.
[60] Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. Ann Intern Med. 2001;134:663-94.
[61] Vandenbroucke JP, von Elm E, Altman DG, Gotzsche PC, Mulrow CD, Pocock SJ, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. Int J Surg. 2014;12:1500-24.
[62] Cohen JF, Korevaar DA, Altman DG, Bruns DE, Gatsonis CA, Hooft L, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. BMJ Open. 2016;6:e012799.
[63] Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. Ann Intern Med. 2009;151:W65-94.

Appendix

A - Methods

Defining the framework structure

In a previous systematic survey[25], we had identified a range of quality definitions, criteria, and themes across different stakeholder groups, as well as a range of quality assessment tools and checklists. These definitions, criteria and themes as well as the derived tools and checklist were consolidated into a comprehensive framework matrix applying the framework method according to Gale [36]. This included the following two- step procedure:
First, we derived the following three structural building blocks for an initial framework (Table 1):

a) quality **dimensions**, with a dimension being defined as an overarching concept of quality containing multiple individual quality questions,
b) successive **study stages** to which the dimensions apply, with a stage being defined as a well-defined period within the continuum of a study, and
c) quality **promoters**, with a promotor being defined as set of factors that may enhance all listed quality dimensions at a research institution.

Table 1. Initial framework matrix

	Research stage 1	**Research stage 2**	**Research stage …**
Quality dimension 1	- Item 1 - Item 2 - Item 3	…	
Quality dimension 2	…	…	
Quality dimension …			

Second, quality definitions, criteria, or themes acquired through our systematic search were first coded into quality items (i.e. single aspects of quality) and then thematically grouped into overarching quality dimensions. We identified a total of six quality dimensions and five successive temporal research stages, resulting in a 5x6 matrix. Two groups of items, those belonging to infrastructural aspects and the sustainability aspect of educating junior researchers, did not fit within one dimension or temporal stage and were included as quality promoters. We subjected this initial framework to iterative consultation about comprehensiveness and subsequent editing by the authors and affiliated interested academics until we reached internal consensus.

Delphi process

We subjected the framework to a modified online Delphi process consisting of three successive stages:

i) Identification and invitation of stakeholder representatives

ii) Delphi-rounds 1 and 2: Identification of any additional quality item that we had not yet considered, and establishing broad consensus across stakeholders on the overall framework structure

iii) Delphi rounds 3 and 4: Seeking agreement on a more refined framework including main quality questions and descriptive examples, with a focus on operationalization in the Swiss academic setting

i) Identification of stakeholder representatives

To allow for broad inclusion of perspectives, we considered the same seven stakeholder groups to be relevant as in the systematic survey [25]: (1) patient organizations and representatives, (2) academic national research institutions/initiatives, clinical investigators, academic clinical trial units, methodological researchers (3) national and supranational governmental bodies, (4) regulatory agencies, (5) ethics committees, (6) the pharmaceutical industry and contract research organizations, and (7) funding agencies.

Our team, with help from affiliated collaborators, and by word of mouth among the related networks (e.g. European Patient Academy on Therapeutic Innovation (EUPATI) for patient representatives) identified potential stakeholder representatives from 16 countries. We recruited participants on the basis of awareness of quality issues related to clinical research and ability to provide feedback within a specified time window.

ii) Delphi-rounds 1 and 2

In round one and two, 109 survey participants from 16 countries were invited through the survey software SurveyMonkey© (www.surveymonkey.net) to provide their comments on the overall suitability and the comprehensiveness of the proposed framework structure, and the individual items to be included. These two Delphi-rounds aimed at (i) identifying any additional quality item that we had not yet considered, and (ii) establishing broad consensus across stakeholders on the overall framework structure. Consensus was pre-defined as an agreement of 80% or higher. Only stakeholders who responded in round one were invited to participate in the following round. After each round, we shared with respondents a summary of the adaptations made based on their suggestions in the previous round and asked for their agreement or further improvements and suggestions on structure and content. Of the 109 invited, 58 (53%) participants provided suggestions or comments in the first round; and

45/109 completed both rounds (Table 1). In each round, we sent two reminders via SurveyMonkey©.

Table 1. **Response rates by stakeholder group for Delphi rounds one and two**

		Round 1		Round 2	
	Number of participants invited (thereof (Swiss)	**No of respondents (thereof Swiss)**	**Total response rate,% [1]**	**No of respondents (thereof Swiss)**	**Total response rate,% [1]**
Stakeholders					
Patient group / representative	20 (19)	10 (10)	50.0	7 (7)	70
Academia	39 (11)	28 (8)	71.8	23 (8)	82.1
Pharmaceutical Industry / CROs	12 (5)	10 (5)	83.3	6 (4)	60
Ethics committee / IRB	8(5)	4 (3)	50.0	3 (2)	75
Governmental bodies & Regulatory bodies	24(2)	3 (1)	12.5	3 (1)	100
Funding Agency	6 (2)	3 (2)	50.0	3 (2)	100
No of countries[2]	16	13		11	
Total	**109**	**58**	**53**	**45**	**78**

[1] Response rates were calculated based on the number of respondents in each round compared to the respondents in the previous round; only respondents were invited to participate in further rounds of the survey

[2] International organizations or companies: Location of headquarters

iii) Delphi-rounds 3 and 4

Seeking consensus on how to operationalize the framework structure in the Swiss academic setting, for Delphi rounds three and four, we invited additional 33 stakeholder representatives from Switzerland, particularly academics (Table 2). In particular, we invited representatives (board members and executive directors) of all six Swiss Clinical Trial Units at University hospitals and members of the executive committee and the Quality Working Group at the Swiss Clinical Trial Organization. For this round, the previous “quality items” were rephrased as “main quality question” accompanied by descriptive examples in order to allow operationalization of the framework (Table 3). We asked for the agreement (yes/no) on the adapted framework structure, content, and wording of main quality questions and corresponding examples and allowed for free text comments on the suitability, the comprehensiveness, and the completeness of dimensions and items for each research stage. In round four, we additionally provided respondents with all anonymized comments, a response by the authors to each comment, and the overall agreement score on framework structure and main quality questions Participants were again asked for their agreement on

structure and content of the framework and were allowed to suggest specific adaptations to the framework using a shareable, but anonymized, googledocs.com (https://docs.google.com) format. Final adaptations to the framework were made by the authors through iterative discussion and shared with the Delphi participants. After round four, an agreement of over 80% was reached for the structure as well as the main quality questions in each research stage (Table 2). In each round, we sent minimum two email reminders.

Table 2. Response rates by stakeholder group for Delphi rounds three and four

	Round 3					Round 4	
	No of participants invited from round 2 (thereof Swiss)[1]	**No of respondents (thereof Swiss)**	**No addit. invited participants[2] (No of respondents)**	**Total No of respondents (thereof Swiss)**	**Total response rate, % (thereof Swiss)[3]**	**No of respondents[4] (thereof Swiss)**	**Response rate, % (thereof Swiss)**
Stakeholders							
Patient group / representative	7 (7)	3 (3)	0 (0)	3 (3)	42.8 (100)	3 (3)	100 (100)
Academic representatives:	23 (8)	14 (5)	28 (19)	33 (24)	64.7 (66.7)	31 (24)	93.9 (100)
National research institutions	6 (4)	5 (4)	9 (4)	9 (7)	60.0 (53.8)	9 (7)	100 (100)
Clinical investigators	4 (2)	2 (1)	5 (4)	7 (6)	77.8 (85.7)	7 (6)	100 (100)
Academic Clinical Trial Units	0 (0)	0 (0)	11 (9)	9 (9)	81.8 (81.8)	9 (9)	100 (100)
Methodological research	13 (2)	7 (2)	3 (2)	9 (2)	56.3 (40)	7 (2)	77.8 (100)
Pharmaceutical Industry / CROs	6 (4)	4 (2)	1 (1)	5 (3)	71.4 (60)	5 (3)	100 (100)
Ethics committee / IRB	3 (2)	3 (2)	1 (1)	4 (3)	100 (100)	4 (3)	100 (100)
Governmental bodies & Regulatory bodies	3 (2)	3 (2)	2 (2)	5 (4)	100 (100)	5 (4)	100 (100)
Funding Agency	3 (2)	3 (2)	1 (1)	4 (3)	100 (100)	4 (3)	100 (100)
No of countries	11	9	1	9		7	
Total	**45 (24)**	**30 (16)**	**33 (24)**	**54 (40)**	**69.2 (70.2)**	**52 (40)**	**96.3 (100)**

[1] All respondents from round two were invited to participate in round three
[2] Predominantly representatives of Swiss academia, n=3 were non-Swiss
[3] Response rates were calculated based on the number of participants invited from round two and the additional Swiss participants invited for round three and four only
[4] Participants who responded to round three were invited to participate in round four

B - All versions of the framework including Delphi survey comments.

Available on request.

C - Results

Table 1. Agreement scores of Delphi participants on framework structure and content. Only includes those participants who gave an answer to the respective question.

	Delphi round 3, n (%)	Delphi round 4, n (%)
Overall framework structure[1]	Total : 47/54 (87.0) Swiss only: 36/40 (90.0)	Total: 52/52 (100) Swiss only: 40/40 (100)
Stage I: Conceptualization, main questions	Total: 40/52 (76.9) Swiss only: 30/40 (75.0)	Total: 52/52 (100) Swiss only: 40/40 (100)
Stage II: Planning and feasibility, main questions	Total: 39/51 (76.5) Swiss only: 30/39 (76.9)	Total: 51/52 (98.1) Swiss only:39/40 (97.5)
Stage III: Conduct, main questions	Total: 43/51 (84.3) Swiss only: 32/37 (86.5)	Total: 52/52 (100) Swiss only: 40/40 (100)
Stage IV: Analysis and Interpretation, main questions	Total: 43/51 (84.3) Swiss only: 32/37 (86.5)	Total: 52/52 (100) Swiss only: 40/40 (100)
Stage V: Reporting and dissemination, main questions	Total: 41/ 51 (80.4) Swiss only: 31/37 (83.8)	Total: 52/52 (100) Swiss only: 40/40 (100)
Quality promoter: Infrastructure	Total: 45/51 (88.2) Swiss only: 34/37 (91.9)	Total: 52/52 (100) Swiss only: 40/40 (100)
Quality promoter: Education and sustainability	Total: 44/51 (86.3) Swiss only: 35/37 (94.6)	Total: 50/52 (96.2) Swiss only: 38/40 (95.0)

[1]Agreement score on overall framework structure in round 1: 26/49 (53.1%); in round 2: 33/34 (97.1%).

Detailed description of quality dimensions

Protection of patients' safety and rights

The first quality dimension, *protection of patients' safety and rights*, represents the cornerstone of research and is therefore a conditio-sine-qua-non dimension in our framework. It assures that participants' safety, rights, and well-being are respected and protected at all times.

Participants should be informed about the research and provide their voluntary consent, but also have the opportunity to withdraw.[1-3] Within the context of standard clinical practice and the research protocol, potential benefits to individuals and the society must outweigh the risks[4],[5] and there should be clinical equipoise- the absence of a consensus regarding the comparative merits of the interventions to be tested.[1, 6] During and after the conduct of the study, participants' rights, safety, and privacy must be protected at all times. Further, study participants should be selected in a fair and equitable manner.[1] Moreover, the research

protocol must be practically feasible. For example, research that could not possibly enroll sufficient participants cannot generate valid scientific knowledge and is thus unethical.[1, 4] Feasibility assessments prior to study start are crucial in order to avoid waste in financial and human resources – and to justify exposing participants to burdens or risks - leading to the many clinical trials that are terminated prematurely.[7-15]

Relevance, patient centeredness and involvement

The second quality dimension, *research relevance and patient centeredness and involvement* reflects the extent to which the research question is scientifically and societally beneficial (i.e. leads to improved decision-making in health care) and involves patient values and preferences at all stages.

Relevant research should build on what is already known, preceded by systematic reviews.[16, 17] Further, it should address a question leading to clinically relevant information gain [18, 19] avoiding subjective approaches that may be unduly influenced by special interest groups.[20] Institutions should reward rigorous replication of previous work in order to battle the low rate of confirmation.[21] Ideally, this is incorporated upfront in designing the research agenda in a given field in order to avoid multiple necessary replications or redundant meta-analyses combining them.[19, 22] Further, relevant research is patient centered and should be aligned with patient priorities, the utilities patients assign to different problems and outcomes, and how acceptable they find interventions over the period for which they are indicated.[7, 23, 24] As suggested by initiatives such as the Patient-Centered Outcomes Research Institute (PCORI) [25], the James Lind Alliance (www.lindalliance.org), or the INVOLVE Initiative in the UK (www.invo.org.uk), patient values and preferences should be fostered through patient (representative) engagement during all stages of a clinical study. For example, through close collaboration with patient organizations at all stages, adaptations of inclusion and exclusion criteria where necessary, or appropriate dissemination of lay language summaries of study outcomes.

Minimization of bias – Internal validity

The third quality dimension, *minimization of bias – or internal validity*, reflects the extent to which systematic error (bias) is minimized, i.e. through selecting an appropriate study design and pre-specifying analyses.

Minimizing bias and thereby maximizing internal validity is dependent on the chosen study design and has been described to be difficult to avoid.[26] Established tools such as the Cochrane Risk of Bias tool for RCTs [27], ROBINS-I for observational research [28] or

QUADAS-2 for diagnostic accuracy studies [29] provide guidance on how to plan and conduct studies with minimal bias. An effective solution to mitigate self-deception, for example, is blinding, which is applicable to some research contexts. Chosen outcomes should be pre-specified, valid, reliable, measured at appropriate times, and comparable across similar trials.[30-32] Data collection should then be conducted in accordance with the procedures pre-specified in the protocol. In general, the adoption of appropriate statistical methods [33], standardized definitions and analyses and stringent thresholds for claiming discoveries success [34] may decrease false-positive rates. Finally, conflicts of interests should be avoided, or at least transparently reported, in order to avoid spinning of more favorable conclusions due to the involvement of conflicted parties.[35, 36]

Precision

The fourth quality dimension, *precision - or statistical validity*, reflects the extent to which random error is minimized (i.e. sufficiently narrow confidence intervals are achieved to confirm or reject clinical hypothesis), and to what level precision is reported and described in order for readers to be able to judge it.

The development and approval of valid study methods and improvements in study design have been described to improve the precision, and therefore reliability of results.[37] Then, efforts need to be made to minimize variability of study procedures and measurement error throughout study conduct to guarantee interpretable data and an ethical study conduct.[1]
It is further important that expected treatment effects and event rates in intervention and control groups are realistic, and that estimates are based on empirical evidence. Validated, non-surrogate outcomes should provide clinical insights to claim power.[7, 38-40] Sample sizes should be justified to measure the expected impact, and recruitment should continuously be monitored to ensure successful reach of target sample size.

Transparency and access to data

The fifth quality dimension, *transparency and access to data*, reflects the extent to which study planning, conduct, data collection and presentation of results are transparent to and accessible for the scientific community and the public. It includes the registration of the study in a publicly accessible database, publication of the full study protocol, publication of the study results - independent of their effect size or direction-, and explicitly, encouraging access to the full patient-level data set (data sharing).

Reporting, review, publication, dissemination, and post-publication review of research shape its reliability.[35] There are currently over 300 reporting guidelines to improve and

standardize reporting (e.g. as catalogued by the EQUATOR Network, http://www.equator-network.org/) and multiple ideas about how to change dissemination of information.[41] Yet, studies that obtain positive and novel results are more likely to be published than studies that obtain negative results or report replications of prior results.[26, 41-43] Research should be pre-registered, as promoted by websites such as Open Science Framework (http://osf.io/) in order to enhance transparency.[44-47] Registration has been proposed for many types of research, including observational studies. Reporting of outcomes should be completely consistent with the pre-registered commitments, and avoid adding new ones (see www.COMPare-trials.org).

Further, sharing of data, protocols, materials, and softwares should be promoted as happening in several –omics fields and may similarly improve the credibility and reproducibility of clinical research studies.[26, 48] The TOP guidelines[67] promote open practices while an increasing number of journals and funders require open practices (for example, open data), with some offering their researchers free and open access publication.

Generalizability – external validity

The sixth quality dimension, *generalizability – or external validity-*, reflects the extent to which study results are applicable and generalizable to the wider patient population in real life circumstances.

Thus, the characteristics of planned study participants should be representative of patients who would use the intervention after study end. Further, the flow of participants through each stage including the reasons for which patients left the study before its end should be documented and reported, together with the results of pre-specified subgroup analyses of key patient characteristics (e.g. disease severity, age or gender). Treatment effects may be similar in nonparticipants and capturing real-life circumstances is possible by utilizing pragmatic study designs [49] allowing for greater applicability and consideration of external validity.[50-52] In 2009, a first tool called the Pragmatic Explanatory Continuum Index Summary (PRECIS) was published to help researchers think more carefully about the impact their design decisions would have on applicability.[53] In 2015, an improved, validated version of PRECIS was published providing guidance on how to match design decision to how the trial results are intended to be used.[51]

Description of quality promoters

Sustainability and Education

Examples of good scientific conduct should be used in practice to train early-career researchers, making quality sustainable. In addition, proper training and continuing education of scientists in research methods and statistical literacy are important to train physicians in critical thinking skills instead and evidence-based research instead of simply producing more papers.[54] Common statistical misperceptions and interpretations could be addressed through improved training.[24, 26] Moreover, methodological best practices are under constant revision and improvement so that senior as well as junior researchers need continuing education, not least because much training of early-career researchers is informal and flows from their supervisors or mentors.[26] Educational resources should be accessible, easy-to-digest and immediately and effectively applicable to research in order to maximize their use.[26]

Infrastructure

In addition to infrastructural support such as space, equipment, or materials, the need for independent methodological support is well established, particularly for clinical trials. Many of them have multidisciplinary steering committees to provide advice and oversee the design and conduct of the trial. Including independent experts in the design, monitoring, analysis or interpretation of research outcomes may not only improve the study, but also mitigate influences such as financial or non-financial conflicts of interests of the investigators.[55, 56] Collaboration across many study sites can- instead of relying on the limited resources of single investigators- facilitate high-powered designs, standardization, and greater potential for testing generalizability across the settings and populations sampled.[26]

References

[1] Emanuel EJ, Wendler D, Grady C. What makes clinical research ethical? JAMA. 2000;283:2701-11.
[2] The Ethics of Consent: Theory and Practice. New York: Oxford University Press; 2009.
[3] Miller FG, Wertheimer A. The fair transaction model of informed consent: an alternative to autonomous authorization. Kennedy Inst Ethics J. 2011;21:201-18.
[4] Laurence DR. Ethics and Regulation of Clinical Research. J Med Ethics. 1988;14:44-6.
[5] World Medical Association declaration of Helsinki. Recommendations guiding physicians in biomedical research involving human subjects. Jama. 1997;277:925-6.
[6] Freedman B. Equipoise and the ethics of clinical research. N Engl J Med. 1987;317:141-5.
[7] Ioannidis JP. Why Most Clinical Research Is Not Useful. PLoS Med. 2016;13:e1002049.
[8] Kasenda B, von Elm E, You J, Blumle A, Tomonaga Y, Saccilotto R, et al. Prevalence, characteristics, and publication of discontinued randomized trials. JAMA. 2014;311:1045-51.
[9] Blumle A, Schandelmaier S, Oeller P, Kasenda B, Briel M, von Elm E. Premature Discontinuation of Prospective Clinical Studies Approved by a Research Ethics Committee - A Comparison of Randomised and Non-Randomised Studies. PLoS ONE. 2016;11:e0165605.
[10] Briel M, Olu KK, von Elm E, Kasenda B, Alturki R, Agarwal A, et al. A systematic review of discontinued trials suggested that most reasons for recruitment failure were preventable. J Clin Epidemiol. 2016;80:8-15.
[11] Kasenda B, von Elm E, You J, Blumle A, Tomonaga Y, Saccilotto R, et al. Prevalence, characteristics, and publication of discontinued randomized trials. Jama. 2014;311.
[12] Rosenthal R, Kasenda B, Dell-Kuster S, von Elm E, You J, Blumle A, et al. Completion and publication rates of randomized controlled trials in surgery: an empirical study. Ann Surg. 2015;262.
[13] Schandelmaier S, Tomonaga Y, Bassler D, Meerpohl JJ, von Elm E, You JJ, et al. Premature Discontinuation of Pediatric Randomized Controlled Trials: A Retrospective Cohort Study. J Pediatr. 2017;184:209-14 e1.
[14] Schandelmaier S, von Elm E, You JJ, Blumle A, Tomonaga Y, Lamontagne F, et al. Premature Discontinuation of Randomized Trials in Critical and Emergency Care: A Retrospective Cohort Study. Crit Care Med. 2016;44:130-7.
[15] Stegert M, Kasenda B, von Elm E, You JJ, Blumle A, Tomonaga Y, et al. An analysis of protocols and publications suggested that most discontinuations of clinical trials were not based on preplanned interim analyses or stopping rules. J Clin Epidemiol. 2016;69:152-60.
[16] Clarke M, Chalmers I. Discussion sections in reports of controlled trials published in general medical journals: islands in search of continents? Jama. 1998;280:280-2.
[17] Clarke M, Hopewell S, Chalmers I. Reports of clinical trials should begin and end with up-to-date systematic reviews of other relevant evidence: a status report. J R Soc Med. 2007;100:187-90.
[18] Evangelou E, Siontis KC, Pfeiffer T, Ioannidis JP. Perceived information gain from randomized trials correlates with publication in high-impact factor journals. J Clin Epidemiol. 2012;65:1274-81.
[19] Chalmers I, Bracken MB, Djulbegovic B, Garattini S, Grant J, Gulmezoglu AM, et al. How to increase value and reduce waste when research priorities are set. Lancet. 2014;383:156-65.
[20] Minelli C, Baio G. Value of Information: A Tool to Improve Research Prioritization and Reduce Waste. PLoS Medicine. 2015;12:e1001882.
[21] Ioannidis JP. Why most published research findings are false. PLoS Med. 2005;2:e124.
[22] Siontis KC, Hernandez-Boussard T, Ioannidis JP. Overlapping meta-analyses on the same topic: survey of published studies. Bmj. 2013;347:f4501.
[23] Mullins CD, Vandigo J, Zheng Z, Wicks P. Patient-centeredness in the design of clinical trials. Value Health. 2014;17:471-5.

[24] Ioannidis JP, Greenland S, Hlatky MA, Khoury MJ, Macleod MR, Moher D, et al. Increasing value and reducing waste in research design, conduct, and analysis. Lancet. 2014;383:166-75.
[25] Selby JV, Lipstein SH. PCORI at 3 years--progress, lessons, and plans. N Engl J Med. 2014;370:592-5.
[26] Munafò MR, Nosek BA, Bishop DVM, Button KS, Chambers CD, Percie du Sert N, et al. A manifesto for reproducible science. Nature Human Behaviour. 2017;1:0021.
[27] Cochrane Risk of Bias Tool.
[28] Sterne JA, Hernan MA, Reeves BC, Savovic J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. Bmj. 2016;355:i4919.
[29] Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155:529-36.
[30] Gorst SL, Gargon E, Clarke M, Smith V, Williamson PR. Choosing Important Health Outcomes for Comparative Effectiveness Research: An Updated Review and Identification of Gaps. PLoS ONE. 2016;11:e0168403.
[31] Williamson P, Clarke M. The COMET (Core Outcome Measures in Effectiveness Trials) Initiative: Its Role in Improving Cochrane Reviews. Cochrane Database Syst Rev. 2012:ED000041.
[32] Williamson PR, Altman DG, Blazeby JM, Clarke M, Devane D, Gargon E, et al. Developing core outcome sets for clinical trials: issues to consider. Trials. 2012;13:132.
[33] Nuzzo R. Scientific method: statistical errors. Nature. 2014;506:150-2.
[34] Johnson VE. Revised standards for statistical evidence. Proc Natl Acad Sci U S A. 2013;110:19313-7.
[35] Ioannidis JP. How to make more published research true. PLoS Med. 2014;11:e1001747.
[36] Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, Julious S, et al. Reducing waste from incomplete or unusable reports of biomedical research. Lancet. 2014;383:267-76.
[37] Poste G. Biospecimens, biomarkers, and burgeoning data: the imperative for more rigorous research standards. Trends Mol Med. 2012;18:717-22.
[38] Hopewell S, Dutton S, Yu LM, Chan AW, Altman DG. The quality of reports of randomised trials in 2000 and 2006: comparative study of articles indexed in PubMed. Bmj. 2010;340:c723.
[39] Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? Ann Intern Med. 1996;125:605-13.
[40] Ferreira-Gonzalez I, Busse JW, Heels-Ansdell D, Montori VM, Akl EA, Bryant DM, et al. Problems with use of composite end points in cardiovascular trials: systematic review of randomised controlled trials. Bmj. 2007;334:786.
[41] Chan AW, Song F, Vickers A, Jefferson T, Dickersin K, Gotzsche PC, et al. Increasing value and reducing waste: addressing inaccessible research. Lancet. 2014;383:257-66.
[42] Song F, Parekh S, Hooper L, Loke YK, Ryder J, Sutton AJ, et al. Dissemination and publication of research findings: an updated review of related biases. Health Technol Assess. 2010;14:iii, ix-xi, 1-193.
[43] Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. Lancet. 2009;374:86-9.
[44] Tse T, Williams RJ, Zarin DA. Update on Registration of Clinical Trials in ClinicalTrials.gov. Chest. 2009;136:304-5.
[45] Zarin DA, Tse T. Trust but verify: trial registration and determining fidelity to the protocol. Ann Intern Med. 2013;159:65-7.
[46] Zarin DA, Tse T, Williams RJ, Califf RM, Ide NC. The ClinicalTrials.gov results database--update and key issues. N Engl J Med. 2011;364:852-60.
[47] Zarin DA, Tse T, Williams RJ, Rajakannan T. Update on Trial Registration 11 Years after the ICMJE Policy Was Established. N Engl J Med. 2017;376:383-91.
[48] Doshi P, Goodman SN, Ioannidis JP. Raw data from clinical trials: within reach? Trends Pharmacol Sci. 2013;34:645-7.

[49] Tunis SR, Stryer DB, Clancy CM. Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. JAMA. 2003;290:1624-32.
[50] Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?". Lancet. 2005;365:82-93.
[51] Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. The PRECIS-2 tool: designing trials that are fit for purpose. Bmj. 2015;350:h2147.
[52] Treweek S, Zwarenstein M. Making trials matter: pragmatic and explanatory trials and the problem of applicability. Trials. 2009;10:37.
[53] Thorpe KE, Zwarenstein M, Oxman AD, Treweek S, Furberg CD, Altman DG, et al. A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers. J Clin Epidemiol. 2009;62:464-75.
[54] Collins FS, Tabak LA. Policy: NIH plans to enhance reproducibility. Nature. 2014;505:612-3.
[55] Panagiotou OA, Ioannidis JP. Primary study authors of significant studies are more likely to believe that a strong association exists in a heterogeneous meta-analysis compared with methodologists. J Clin Epidemiol. 2012;65:740-7.
[56] Riaz H, Raza S, Khan MS, Riaz IB, Krasuski RA. Impact of Funding Source on Clinical Trial Results Including Cardiovascular Outcome Trials. Am J Cardiol. 2015;116:1944-7.

D - Full framework, including main quality questions and examples

Study Stage I: Concept **Milestone: Research question including study type defined and viable**		
Dimension	**Main question**	**Examples**
Protection of patient safety & rights	Can the research question be answered in the given setting?	Based on a rough resource assessment, and potentially available study participants, is it feasible to answer the research question?
		Based on a rough budget estimate, is it feasible to answer the research question with a specified study type?
	Does study consider equity appropriately?	Are participants selected so that :
		vulnerable individuals are neither targeted for risky research nor withheld from research relevant to these populations?
		socially powerful individuals are not favored for potentially beneficial research?
	Is the research design adequate for the stage of an investigated technology to ensure patient safety?	Are sufficient data on toxicity/teratogenicity of an intervention available from animal studies or phase I studies?
	Do the (assumed) short and long term benefits of the study outweigh potential risks associated with the study (consistent with clinical equipoise)?	
Relevance / Patient centeredness & involvement	Is significant add-on value to already existing evidence given, taking into consideration burden of disease and anticipated benefit of treatment?	Are uncertainties in existing evidence identified and discussed in a systematic review?
		Does research:
		Expand or challenge current knowledge?
		Open additional areas for new research activity?
		Justify replication of existing evidence, if applicable?
	Are patient representatives/ advocates and their needs and values adequately involved in the development of the research question?	
	Are outcome measures patient-relevant?	Are outcomes patient-relevant, including quality of life, if applicable, and with judicious use of surrogate endpoints?
Minimization of bias (internal validity)	Is the selected study type/design appropriate to minimize bias?	Is the study randomized or, if not, appropriately controlled for confounding?
	Are potential sources of bias anticipated, evaluating the magnitude and the likely direction?	
	Are outcome measures well-defined, pre-specified, valid, reliable and measured at appropriate times?	Are outcomes: well-defined (upfront)?
		valid (measure what they intend to measure)?
		reliable(stable and consistent when repeatedly measured)?
		sensitive to important change?
		measured at appropriate times?
		standardized across studies (core outcome sets, if applicable)
Precision	Has estimate of the required sample size been made (for feasibility purposes, see "Protection of patient safety & rights")?	
Transparency / Access to data	Is the research question clearly specified (e.g. in a synopsis)?	Is each component of P(I/E)(C)O[1] as applicable to study design clearly defined, i.e. :
		- Patient population to be recruited/investigated in the study
		- Intervention to be assessed,
		- Exposure to be assessed,
		- Diagnostic Test to be assessed,
		- Control intervention as comparator,
		- Outcomes to be measured?
Generalizability (external validity)	Are planned study participants representative of patients who would use the drug/intervention/diagnostic test in a real-life setting?	Are unnecessary restrictions through inclusion/exclusion criteria avoided (to facilitate rapid accrual, broader generalization, pragmatic study conduct)?
		Is the control group adequate given current evidence and clinical practice (e.g. "standard of care" rather than "no treatment")?

Study Stage II: Planning & Feasibility **Milestone: Protocol developed and approved by regulatory bodies**		
Dimension	**Main question**	**Examples**
Protection of patient safety & rights	Do the potential short and long term benefits of the study outweigh study burden (due to study visits, intervention, procedures etc.)?	
	Are patients' safety and rights protected through the study's adherence to applicable national and international regulations and laws?	Are study documents (e.g. protocol, participant information etc.) written in accordance with applicable national (and international, if applicable) regulations/laws?
		Are informed consent documents written in lay language and easily understandable for study participants?
		Has approval been obtained from ethics committee?
		Has approval been obtained from regulatory agency (if applicable)?
	Has feasibility been checked thoughtfully based on existing evidence?	Is valid and robust preclinical data present (if applicable)?
		Have crucial feasibility aspects (e.g. recruitment) been piloted?
		Are recruitment assumptions realistic in a specified timeframe (e.g. empirical data from electronic health records or from pilot study present)?
		Have national/ international study registries been checked for studies that could interfere with the planned study?
		Do anticipated study costs (preparation, conduct, analysis, dissemination) match with available budget?
		Is study cost data related to planning, conduct, analysis, and dissemination planned to be collected (if applicable)?
	Is collection, documentation, and reporting of Adverse Events / Serious Adverse Events / Suspected Unexpected Serious Adverse Reaction according to the applicable regulations planned and specified in the protocol?	
	Are mechanisms established which allow early study termination when required and prevent early study termination for inadequate reasons?	Is one or few interim analyses for safety considered?
		Is early stopping for benefit with insufficient collection of safety data avoided [9]?
Relevance / Patient centeredness & involvement	Is knowledge transfer/use (e.g. plans for inclusion of results in clinical guidelines) planned?	Are relevant guideline groups identified and contact established?
		Are patient representatives involved in protocol development?
Minimization of bias (internal validity)	Is statistical analysis pre-specified (using outcomes as defined in concept stage)?	Are outcomes, datasets, subgroups, handling of missing data, etc., pre-specified?
	Is study monitoring (adapted to risk of study, if applicable) planned and documented in a monitoring plan??	
	Is data management planned and documented in a data management plan?	
	Is minimization of bias planned for according to the research question and study design?	Exemplary items according to study type (non-exhaustive):
		Randomized Controlled Trials:
		Please also refer to Cochrane Risk of Bias tool for RCTs [1] for full list of items.
		Is randomization adequate and concealed?
		Are (known) prognostic factors distributed equally (i.e. are groups prognostically balanced at the start of the trial)?
		Is blinding of participants and/or care-givers adequate?
		Are concomitant interventions documented?
		Is blinding of outcome assessors adequate?
		Are plans to minimize losses to follow up present?

		Are plans to analyze study participants in groups as randomized present?
		Observational studies (incl. cohort studies):
		Please also refer to ROBINS-I tool [2] for full list of items.
		Is collection of data carefully planned, i.e. are all relevant confounders considered and measured?
		Are all study participants selected or recruited from the same or similar populations (incl. the same time period)?
		Do the study participants represent the cases originated in the community? (e.g. due to issues with healthcare access)
		Are inclusion and exclusion criteria pre-specified and applied uniformly to all study participants?
		Are plans to minimize losses to follow-up present?
		Is timeframe sufficient so that one can reasonably expect to see an association between exposure and outcome if it existed?
		For exposures that can vary in amount or level, does the study examine different levels of the exposure as related to the outcome (e.g. categories, or exposure measured as continuous variable)?
		Is exposure measured more than once over time?
		Diagnostic accuracy studies:
		Please also refer to QUADAS-2 Risk of Bias tool [3] for full list of items.
		Is there an independent, blind comparison between index test and an appropriate gold standard of diagnosis?
		Is the diagnostic test evaluated in a representative, and ideally full spectrum of study participants (like those in whom it would be used in practice, spectrum ranging from mild to severe, and early to late cases of target disorder)?
		Is a reference standard applied regardless of the index test results (ideally both index test and reference standard should be carried out on all study participants)?
		If no, is it planned to follow up study participants for an appropriate period of time (dependent on disease in question) to see if they are truly negative?
Precision	Are expected treatment effects and event rates in intervention and control groups realistic and estimated based on empirical evidence?	Is sample size realistically estimated and clearly described (incl. assumed treatment effects, references for estimates, power, alpha error, and expected losses to follow-up)?
		Is consent rate precisely estimated?
		Are treatment effects and/or event rates estimated in both intervention and control groups?
		If yes, are they based on evidence such as systematic literature reviews, meta-analysis?
		Is rationale for non-inferiority / equivalence design provided (if applicable)?
		Is rationale for maximum clinically acceptable difference (equivalence margins) provided (if applicable)?
		Is rationale for sample size given if not derived statistically?
	Are recruitment procedures and recruitment monitoring planned to ensure sufficient sample size?	
Transparency / Access to data	Is the protocol in accordance with established standards (e.g. SPIRIT [4] or other applicable guidelines depending on study design)?	Is protocol peer-reviewed?
		Is publication and accessibility of full study protocol planned?
		Is study registered in publicly accessible database / registry?
		Does protocol state a plan on how to deal with study publication in case target sample size could not be achieved/study had to be discontinued prematurely?
	Is there a dissemination plan to share study information including the protocol, summary results, and participant level data?	

Generalizability (external validity)	Are study procedures/observations in line with routine practice in the given setting?	Is standard of care/current practice clearly defined?
		Are interventions /observations close to foreseen everyday practice?
		Is participant follow up close to everyday practice?

Study Stage III: Conduct **Milestone: Last patient last visit**		
Dimension	**Main question**	**Examples**
Protection of patient safety & rights	Is respect for and consideration of patient rights, well-being and dignity guaranteed throughout conduct of study?	Are study participants respected at all times, i.e.:
		Is withdrawal from study at any time explicitly permitted?
		Are study participants informed about purpose of research, its procedures (including study medication, concomitant medication, emergency management, etc.) and potential risks, benefits and alternatives, so that they can make a voluntary decision?
		In case of routinely collected data (including biological material), are study participants informed about the further use of their data for research purposes?
		Is study participants' privacy and confidentiality ensured during (and after) study, e.g. through appropriate coding?
	Is patient safety guaranteed throughout conduct of study?	Are study participants informed of newly discovered risks?
		Are side effects / Adverse Events/ Serious Adverse Events/ Suspected Unexpected Serious Adverse Reactions etc. monitored and reported to the ethics committee within required timeframes?
	Is study conducted according to protocol?	
	Is compliance of participants and study staff with protocol monitored?	
	Are patients' safety and rights protected through the study's adherence to applicable national and international regulations and laws?	
Relevance / Patient centeredness & involvement	Are there any measures in place to assure study participants' involvement, cooperation, and feedback throughout conduct of study (e.g. incentives, phone calls, etc.)?	
Minimization of bias (internal validity)	Are data systematically collected as pre-specified in protocol?	Are losses to follow-up minimized?
		Are protocol deviations documented, and reported to the respective institutions?
		Are changes in study procedures amended in the protocol?
		Are missing data documented by individual outcomes?
		Apart from the allocated treatment, are study groups treated equally (e.g. no additional treatments or tests)?
		If applicable, are study participants and clinicians kept "blind" to which treatment was being received?
	Is monitoring conducted according to the pre-specified monitoring plan?	
Precision	Is enrollment of study participants monitored?	Are formal techniques in place to monitor recruitment centrally and at participating sites?
		Are measures in place to allow timely reaction in case recruitment deviates from expectations?
	Is variability of study procedures and measurement error minimized, e.g. by utilizing centralized monitoring strategies?	
Transparency / Access to data	Is study conduct transparent to all involved parties?	Are protocol amendments or any necessary deviations from the original protocol clearly documented and disseminated to appropriate parties within reporting timelines?
		Are internal or external audits planned, conducted and reported?
		Is an external and independent Data Monitoring Committee present or reason provided, why it is not needed?

Generalizability (external validity)	Are numbers of participants through different stages of a study documented (patient flow) including reasons for leaving the studybefore its end (if voluntarily provided by patient)?	Is proportion of study participants who declined randomization documented?
		Are the reasons for participants leaving the study before its scheduled end documented (if voluntarily provided by patient)?

Study Stage IV: Analysis & Interpretation **Milestone: Study data analyzed and interpreted**		
Dimension	**Main question**	**Examples**
Protection of patient safety & rights	Does data sharing adhere to appropriate data protection policies?	Is patient-level data anonymized?
		Have other risks for re-identifying participants been minimized?
Relevance / Patient centeredness	Are data analyzed so that the use of results by different stakeholders is maximized?	Are confidence intervals calculated on an absolute scale to gauge the benefit of an intervention for decision makers (e.g. clinicians, patients, policy makers)?
Minimization of bias (internal validity)	Is the data analyzed as pre-specified in the protocol/statistical analysis plan?	Are post-hoc analyses clearly labelled as such or as exploratory analyses?
		Is data analysis performed using standard, generally accepted software?
		Are data assumptions checked (e.g. normal distribution) as appropriate for planned statistical tests/modelling?
	Are key confounding variables adjusted for in the analysis (e.g. multivariable analysis), if applicable?	
	Does the analysis follow an adequate strategy to deal with participants in whom treatment or follow-up was not in accordance with study protocol?	Is the intention-to-treat principle followed (i.e. all study participants with data analyzed in groups as randomized) in case of a superiority hypothesis?
		Are both a per-protocol and an analysis following the intention-to-treat principle conducted in case of a non-inferiority hypothesis?
	Are results interpreted with least possible "spin"? (e.g. without intentionally implying greater or lesser effects than actually shown by the data)?	
Precision	Is the uncertainty of results considered in the analysis?	Are confidence intervals or other measures of uncertainty calculated?
		Are reasonable sensitivity analyses for missing data conducted?
		Does interpretation adequately reflect uncertainty?
Transparency / Access to data	Is the analysis code clearly documented and the analysis process reproducible?	
	Are deviations from the statistical analysis plan or protocol adequately documented and reported?	
Generalizability (external validity)	Does the interpretation put the results adequately into context of clinical practice/public health?	

Study Stage V: Reporting & Dissemination **Milestone: Study archived and published**		
Dimension	**Main question**	**Examples**
Protection of patient safety & rights	Is study completion/termination communicated to appropriate parties and documented in registries?	Is study completion/termination reported to ethics committee/regulatory bodies?
		Is study completion/termination appropriately documented in national/international registry?
	Are study participants informed about outcome/main findings of the study in plain language (including treatment allocation of participant, if applicable)?	
	Do study participants get access to products/interventions after study, if applicable?	

Relevance / Patient centeredness & involvement	Do authors critically reflect on research findings (results as well as challenges or mistakes during study conduct) and the implications for future research?	
	Is the study easily available to decision/policy/guideline makers?	Is the study cited in a clinical guideline?
	Are study patients/patient representatives involved in the reporting of the study?	Are patient representatives involved in reporting of the study, e.g. in writing of lay term summaries?
Minimization of bias (internal validity)	Are all outcomes and important study characteristics reported as pre-specified in the protocol (outcome reporting bias prevented)?	Are all patient-relevant outcomes reported as pre-specified in the protocol?
		Are important modifications to the protocol (e.g. premature discontinuation) reported (if applicable)?
Precision	Are absolute and relative treatment effects reported accompanied by confidence intervals?	
	Is the analysis set of participants clearly specified?	Are the actual numbers of recruited, randomized (if applicable), followed-up, and analyzed participants reported for each outcome and for each treatment group (if applicable)?
Transparency / Access to data	Is dissemination of data and study results maximized?	Is dissemination maximized through open access?
		Is anonymized individual participant-level data made available (data sharing)?
		Are study results posted in study registries?
		Does publication in journals include full protocol and statistical analysis plan?
		Is dissemination maximized through use of alternative media other than medical journals?
		Are resulting doctoral/master theses made publicly available (if applicable)?
	Are reporting guidelines followed to facilitate critical appraisal and reproducibility?	*Is reference made to reporting guidelines such as CONSORT (Randomised trials) [5], STROBE (Observational studies) [6], STARD (Diagnostic studies) [7], or PRISMA (Systematic reviews) [8] depending on the respective study design.*
		Are detailed methods disclosed in publications (to enable reproducibility)?
	Are selective reporting, "spin", plagiarism and self-plagiarism avoided and conflicts of interest declared?	Is selective reporting of study results avoided?
		Is plagiarism and self-plagiarism avoided?
		Are study results independently peer reviewed?
		Is "spin" (i.e. reporting to convince readers that the beneficial effect of the experimental treatment is greater than shown by the results) minimized in reporting of results?
		Were conflicts of interest declared?
	Is knowledge transfer & exchange fostered?	Is knowledge transfer & exchange fostered through e.g.:
		Community and provider education and outreach
		Facilitation of two-way communication (lay language) with diverse populations and community groups
		Knowledge transfer & exchange among clinical research groups
	Are study records and data sets kept and archived for the legally required period of time?	
Generalizability (external validity)	Is potential impact on clinical practice / public health outlined in publicly accessible research reports (e.g. journal publication)?	
	Are characteristics of included participants clearly reported?	Are inclusion and exclusion criteria clearly reported?
		Are characteristics of included participants clearly reported?
	Are results of pre-specified subgroup analyses, if applicable, reported in order to assess the importance of key participant characteristics (e.g. disease severity, age or gender)?	

Quality promoter: Sustainability / Education	
Main question	**Examples**
Are doctoral students, junior researchers, clinicians, or patient advocates actively involved in all stages of a clinical study, reliably supervised/mentored by senior researchers, and are their specific contributions	Are doctoral students, junior researchers, clinicians, or patient advocates actively involved in study design, planning, conduct, analysis, interpretation and dissemination of results (e.g. publications, conference presentations, reports, or lay summaries)?
	Are doctoral students, junior researchers, or clinicians actively supervised

acknowledged appropriately?	by senior researchers at all stages of a clinical study?
	Are doctoral students, junior researchers, or clinicians mentored as to career options in clinical research (early career development)?
	Are training options and courses in health research methodology available for principal investigators, staff, and patient advocates?
	Are doctoral students, junior researchers, or clinicians mentored to improve awareness about value of clinical research to patients and society as a whole?
	Are processes continuously adapted and improved to changes, developments, issues, and conditions during research continuum (quality by design)?

Quality promoter: Infrastructure	
Main question	**Examples**
Is a Quality Management System incl. Standard Operating Procedures (SOPs) in place?	Is all staff continuously trained in applicable SOPs?
	Are there measures in place to control whether the existing Quality Management System is followed? (i.e. internal audits)
Are well-trained, experienced, and dedicated principal investigators and study staff present?	Has the principal investigator and/or staff been involved in clinical studies before?
	Is all staff continuously trained in GCP and protocol-related activities, and particularly the informed consent process?
	Is training (e.g. GCP) of each participating investigator and staff member clearly documented?
	Are roles and responsibilities of each participating investigator and staff member clearly documented?
	Are all involved stakeholders well and adequately informed about study procedures and changes?
Are expert epidemiologists/methodologists, statisticians, professional data managers, and/or a logistical support unit involved early-on?	Are epidemiologists/methodological specialists involved in development of protocol?
	Are statisticians involved in development of protocol?
	Are data managers involved in the development of the data management plan and the setup of the data management system?
	Is a logistical support unit involved in study planning and/or conduct, e.g. through regulatory affairs experts, study nurses, or project managers?
Are adequate human, material, and equipment resources available for study conduct?	Is dispense, transport, and storage of investigational medicinal product, if applicable, planned?
	Is availability of study-specific materials, hardware, and facilities planned and secured?
	Is a transparent study budget available and approved by experienced personnel, including costs for experts mentioned above?
	Is funding secured through acquisition of competitive money or through collaboration with e.g. industry partners?
Are adequate facilities ensuring data security and privacy in place (incl. competent and effective IT support to facilitate solutions tailored to specific challenges of individual studies or agreement templates for doctoral students with respect to data privacy and confidentiality)?	Is an electronic database incl. audit trail in place?
	Is participant data coded?
	Is IT support present at site?
Is inter-/multidisciplinary collaboration and involvement in study planning and conduct fostered?	Have all relevant stakeholders been involved in protocol development and conduct? (e.g. investigators at other study sites, etc.)
	Is communication between involved staff, sponsor, contractors, and site fostered?
Are all institutions involved in the study covered by compulsory liability insurance?	
Is an overview of the existing research infrastructure available and accessible to any researchers with a study idea?	

E - Potential applications of the framework at different levels and stakeholder groups

Box 1. Possible levels of framework application	
Level of application	**Description**
Stakeholder group	Stakeholder groups differ in interests and needs and may operationalize framework differently, adapted to their setting, whilst still keeping the comprehensiveness of all dimensions represented in the framework (see Box 2 for detailed examples)
Indication area	Indication and treatment areas differ in scientific, ethical, organizational, and cultural aspects and therefore require different study designs to answer the scientific question. The framework should be applied to indication areas by making use of the flexibility of weighing the individual domains, and main questions
Study design	Depending on the study design chosen, different weights across dimensions may have to be applied in order to obtain a meaningful assessment of quality, e.g. for studies in which randomization is unethical and which therefore score lower in internal validity. The examples should support users in applying the framework content to their study design.
Reearch stage	The framework is meant to cover all research phases, from planning to dissemination. However, certain future applications (e.g. templates supporting study planning) may focus on one of the three temporal stages described in the framework.

Box 2. Examples for application of Framework for quality of Clinical Research, by stakeholder group	
Stakeholder	**Examples for application**
Patient organizations / individual patients interested in research participation	- Development of educational material on “good clinical research” - Development of tool/supporting material assisting in the decision of what clinical study to participate in (“how do I know it is a good clinical trial?”), similar to shared-decision aids - Development of “quality labels” for clinical studies issued by patient organizations, i.e. in large international consortia
(Inter-)national funding agencies	- Longitudinal assessment of individual studies, research projects, e.g. in the context of excellency programs - Extension of current criteria for assessment of research proposals, funding decisions
Academic institutions, medical faculties, clinical trial units	- Development of tools for (longitudinal) assessment of research department, research unit, research groups - Development of tools for (longitudinal) assessment of individual studies, research projects, e.g. for cost-efficient allocation of ressources - Provision of report templates - Tools supporting consultancy of trialists at clinical trial units, over entire study lifecycle
Regulatory agencies, health technology assessment bodies, payers (i.e. health insurances)	- Refinement of own quality assessment criteria - Development of longitudinal assessment of individual studies, research projects, e.g. for registration purposes
Ethics committees	- Refinement of own quality assessment criteria - Development of tools for longitudinal assessment of individual studies and their outcomes
Governmental bodies	- Longitudinal assessment of national research program - Guidance on selection process for excellency programs
Drug development/Pharmaceutical industry	- Holistic approach to quality assessment, of both internal decision-making as well as operational aspects at sites
All stakeholders	- Overarching concept for the development/adaptation of context-specific checklists, or scores for the quality of clinical research - Development of “quality labels” based on theoretical foundation increasing trust in clinical research quality - Creation of publicly accessible “quality registries” - Adaptation and application of framework to study designs other than the traditional study types, e.g. Real-World Evidence

Manuscript III: Systematic review on costs and resource use of randomised clinical trials shows a lack of empirical data

Benjamin Speich[a,†], Belinda von Niederhäusern[b,†], Nadine Schur[c], Lars G. Hemkens[a], Thomas Fürst[d,e], Neera Bhatnagar[f], Reem Alturki[g], Arnav Agarwal[f,h], Benjamin Kasenda[a, i], Christiane Pauli-Magnus[b], Matthias Schwenkglenks[c,j]*, Matthias Briel[a,f]* for the MAking Randomized Trials Affordable (MARTA) Group

[a]Basel Institute for Clinical Epidemiology and Biostatistics, Department of Clinical Research, University of Basel and University Hospital Basel, Switzerland

[b]Clinical Trial Unit, Department of Clinical Research, University of Basel and University Hospital Basel, Switzerland

[c]Institute of Pharmaceutical Medicine, University of Basel, Basel, Switzerland

[d]Department of Epidemiology and Public Health, Swiss Tropical and Public Health Institute, University of Basel, Basel, Switzerland

[e]School of Public Health, Imperial College London, London, United Kingdom

[f]Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, ON, Canada

[g]Multi Organ Transplant Center, King Fahad Specialist Hospital Dammam, P.O. Box 15215, Dammam 31444, Saudi Arabia

[h]School of Medicine, University of Toronto, Toronto, Ontario, Canada

[i]Department of Medical Oncology, University of Basel and University Hospital Basel, Switzerland

[j]Epidemiology, Biostatistics and Prevention Institute, University of Zürich, Zürich, Switzerland

† These authors contributed equally to the work reported here and share first authorship.

* These authors contributed equally to the work reported here and share last authorship.

Status: Under review, Journal of Clinical Epidemiology

Abstract

Objectives: Randomised clinical trials (RCTs) are costly. We aimed to provide a systematic overview of the available evidence on resource use and costs for RCTs to identify lever-points which may make RCTs more cost-effective.

Study Design and Setting: We systematically searched MEDLINE, EMBASE and HealthSTAR from inception until November 30, 2016 without language restrictions. We included any publication reporting empirical data on resource use and costs of RCTs and categorised them depending on whether they reported (i) resource and costs of all aspects of an RCT (including planning, conduct, and reporting); (ii) on several aspects, (iii) on a single aspect (e.g. recruitment); and (iv) on overall costs for RCTs. Median costs of different recruitment strategies were calculated. Other results (i.e. overall costs) were listed descriptively. All cost data were translated into USD 2017.

Results: A total of 56 articles that reported on cost or resource use of RCTs were included. None of the articles provided empirical resource use and cost data for all aspects of an entire RCT. Eight articles presented resource and cost data on multiple aspects (e.g. aggregated cost data of different drug development phases, site specific costs, selected cost components). Thirty-five articles assessed costs of one specific aspect of an RCT (i.e. 30 on recruitment; 5 others). The median costs per recruited patient were USD 409 (range: USD 41-6,990). Overall costs of an RCT, as provided in 16 articles, ranged from USD 43-103,254 per patient, and USD 209,202-494 Mio per RCT. For 12 out of 16 articles (75%) the methodology of gathering these overall estimates remained unclear.

Conclusion: Empirical evidence on resource use and costs of RCTs is sparse. Accessible resource use and cost data are urgently needed to make RCTs and the current research environment more transparent and help improve sustainability.

What is new?

Key Findings

- Despite frequent claims that clinical research costs are on constant rise, empirical evidence on resource use and associated costs for randomised clinical trials (RCTs) is sparse
- Current estimates of overall RCT costs cover a wide range, but the underlying methodology of gathering these estimates often remains unclear
- Most articles focus on the comparison of different recruitment strategies, but none provide a detailed overview of all aspects of resource use and associated costs of an RCT

What this adds to what was known?

- To our knowledge this is the first systematic overview of published estimates on trial costs. There is a lack of published empirical data on RCT costs and the evidence base to support the frequently made claim of increasing costs is elusive

What is the implication and what should change now?

- Empirical evidence on cost items for RCTs, specifically in the academic setting, is urgently needed so that lever-points can be identified to make RCTs more cost-effective and feasible

Introduction

Cost estimates for research and development (R&D) for new drugs in 2009 were reported to range from USD 92 million to USD 884 million per compound [1, 2]. The currently highest estimates, published in 2016 by DiMasi et al., are as high as USD 1.4 billion [3]. Several institutions criticized these estimates to be intransparent and raised scepticism whether the high R&D costs reflect reality or if they are overestimated to justify high drug prices [4-6].

Of total R&D costs for new drugs, the clinical phase accounts for the highest expenses [3, 7]. In contrast to the debate about R&D cost estimates there is wide agreement that randomised clinical trials (RCTs) produce the most reliable evidence about the benefits and harms of clinical interventions, ultimately leading to better care for patients [8, 9]. However, they come at high costs [10]. During the last decades a number of initiatives and regulations were implemented to improve research quality and to increase participant protection [11]. However, the complexity and the administrative burden of RCTs increased too and may raise their overall costs and compromise the motivation of investigators to initiate new trials [8, 12, 13]. Stagnating numbers of RCTs [14] and a substantial proportion of initiated RCTs that are prematurely discontinued due to organisational and recruitment problems [15] risk that more uncertainties about medical treatments will remain unaddressed. Thus, efforts to make clinical trials more cost-efficient and feasible are urgently needed. A pre-requisite for such an optimisation process are empirical cost data and evidence on cost-drivers. Therefore, the aim of this study was to systematically review the currently available evidence on resource use and associated cost data for RCTs.

Methods

Literature search and eligibility criteria

An experienced medical librarian (NB) systematically searched Medline, EMBASE and HealthSTAR via Ovid from their inception until November 30 2016 without any language restriction. We used MeSH terms like “Clinical Trials as Topic” and “Cost Analysis” and text words such as “costs” in combination with “clinical trials” (see Online Appendix for the detailed search strategy). Any publication that reported empirical data on resource use and costs of RCTs were included. We excluded articles that assessed the costs or the cost-effectiveness of the intervention only and studies that analysed the cost of R&D in general without giving any specific cost or resource data for RCTs.

Selection of articles and data extraction

In teams of two, we examined all identified titles and abstracts independently and in duplicate (BS, BvN, NS, LH, TF, RA, AA, CPM, MS, MB). For titles and abstracts that were considered potentially relevant by at least one reviewer, we obtained corresponding full texts that were independently reviewed by two authors (BS, BvN). In case of disagreement, a third reviewer (MB) was consulted and the respective article discussed until agreement was reached. Each eligible articles was assigned to one of the following four categories: (i) articles presenting resource and cost data for all aspects of an entire RCT (including planning, conduct and reporting); (ii) articles presenting resource use or cost data for multiple aspects of an RCT (e.g. including costs or resource use of different trial specific steps); (iii) articles presenting resource use or cost data for only one aspect of an RCT (e.g. recruitment); and (iv) articles that reported overall cost numbers (i.e. cost per patient or cost per RCT). The first three categories were mutually exclusive, but not the fourth (overall costs).

From all included articles the following information was extracted by one reviewer and cross-checked by a second reviewer (BS, BvN): Year of publication and time of conduct, population, experimental intervention, control, primary outcome, cost data (including costs for single trial aspects such as recruitment and overall costs).

Analysis

Articles presenting resource or cost data on multiple aspects were descriptively summarised. For articles covering one aspect of costs, the costs of specific recruitment strategies as well as overall costs of recruitment were merged (median; range). Overall costs of RCTs and costs per patient for a complete RCT were descriptively listed per article. In case an article described the costs of several RCTs, we present the range of those costs. All cost data were converted into USD 2017. For the overall costs of RCTs we present the published cost data as well as costs converted into USD 2017. We converted currencies into USD [16] using the start date of the RCT, or, if unknown, the date of the publication. All costs were converted into USD 2017 using the annual inflation listed by the Organisation for Economic Co-operation and Development [17]. If reported, we used the following date in this order of priority to adjust for inflation: (i) date of currency listed in the publication (e.g. USD 2011); (ii) date when the RCT was started; (iii) date of publication.

Results

Search results

Our systematic search yielded 6650 articles after removing duplicates. We excluded 6326 articles based on titles and abstracts (Figure 1), and 268 based on the full text. We included 56 articles (Table 1): (i) there were no articles that presented detailed empirical resource and cost data for all aspects of one or more RCTs; (ii) eight articles presented resource or cost data on multiple aspects of an RCT (three out of these 8 articles also presented overall RCT costs); (iii) 35 articles presented only one specific resource or cost aspect of RCTs (i.e. 30 recruitment costs, 2 costs of protocol development, 1 adverse event costs, 1 pharmacy costs, 1 costs of Good Clinical practice [GCP] compliance); and (iv) 16 articles presented overall costs.

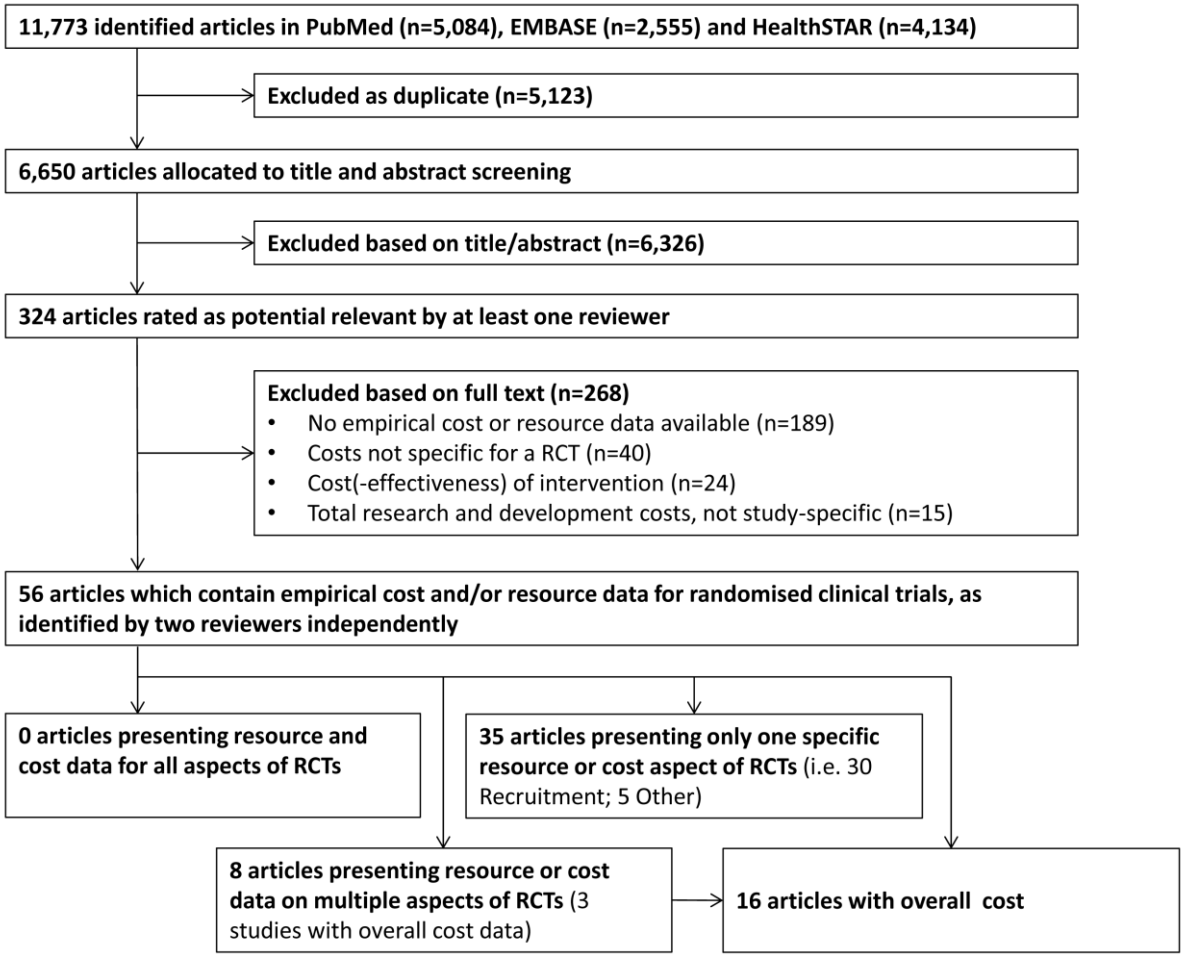


Figure 1: Selection of articles

Table 1: Characteristics of the 56 included articles.

	All articles (n=56)	Articles assessing multiple aspects (n=8*)	Articles assessing only one aspect (n=35)	Overall costs (n=16*)
Year of Publication				
Median Interquartile range Range	2009 2002-2014 1985-2016	2008 1995-2015 1992-2016	2009 2002-2013 1993-2016	2013 1997-2014 1985-2016
Before 2000	10 (17.9%)	2 (25.0%)	5 (14.3%)	4 (25.0%)
2000-2009	19 (33.9%)	2 (25.0%)	14 (40.0%)	3 (18.8%)
2010-2016	27 (48.2%)	4 (50.0%)	16 (45.7%)	9 (56.2%)
Number of trials				
One specific trial	35 (62.5%)	3 (37.5%)	26 (74.3%)	8 (50.0%)
More than one trial	21 (27.5%)	5 (62.5%)	9 (25.7%)	8 (50.0%)
Origin of trials				
North America	33 (58.9%)	4 (50.0%)	25 (71.4%)	6 (37.5%)
Europe	7 (12.5%)	1 (12.5%)	5 (14.3%)	1 (6.3%)
Other	3 (5.4%)	0	3 (8.6%)	0
Unknown	2 (3.6%)	0	0	2 (12.5%)
Multiple	11 (19.6%)	3 (37.5%)	2 (5.7%)	7 (43.8%)
Type of Article				
Original article	51 (91.1%)	7 (87.5%)	34 (97.1%)	13 (81.3%)
Short communication, Editorial/Comment	4 (7.1%)	0	1 (2.9%)	3 (18.8%)
HTA Report	1 (1.8%)	1 (12.5%)	0	0
Medical field				
Medical	18 (32.1%)	2 (25.0%)	10 (28.6%)	7 (43.8%)
Surgery	1 (1.8%)	0	0	1 (6.3%)
Other (e.g. behavioral, physical therapy, psychological therapy, dental treatment)	26 (46.4%)	2 (25.0%)	18 (51.4%)	5 (31.3%)
Multiple trials/unclear	11 (19.6%)	4 (50.0%)	7 (20.0%)	3 (18.8%)
Population				
Adults	32 (57.1%)	4 (50.0%)	20 (57.1%)	10 (62.5%)
Adults and children	3 (5.4%)	0	3 (8.6%)	0
Children	5 (10.7%)	0	3 (8.6%)	2 (12.5%)
Elderly	3 (5.4%)	0	3 (8.6%)	0
Multiple trials	13 (23.2%)	4 (50.0%)	6 (17.1%)	3 (18.8%)

*3 articles presenting overall costs also had data in more detail.

Articles presenting resource and cost data for all aspects of RCTs

We did not find any article that empirically assessed the resource use and associated costs for all aspects of an entire RCT (including planning, conduct and reporting).

Articles presenting resource or cost data for multiple aspects of RCTs

Eight articles presented resource or cost data on multiple aspects for RCTs (Table 2) [18-25]. In short, two studies assessed the overall costs of several industry-sponsored clinical trials stratified by different clinical trial phases [18, 25]. While Sertkaya and colleagues presented

“costs per patient”, “per site cost” and “pre study costs” with additional stratification per medical field, Getz et al. presented costs stratified into “core” and “noncore” procedures. One article from Raftery and colleagues retrospectively assessed the mean costs per patient and other cost components (e.g. costs of statistical input) from studies supported by the British National Health Service (NHS) between 1995 and 2005 [19]. One article authored by Zhu and colleagues [20] listed the overall costs of a large oncology screening RCT and stratified the costs for further screening centres, coordinating centre and for collecting biological samples. Another article assessed time efforts for 41 cancer-trial tasks from 83 Clinical Reseach Associates in 1996 [21]. Thornquist and colleagues estimated the resource use and costs of a large academic prevention trial that was conducted between 1988 and 1991 [23]. While cost components for specific items are listed (e.g. follow up of an active participant for 1 year), resource data (i.e. time efforts) are missing. Two studies published in the 1990s assessed site-specific resources and costs [22, 24].

Table 2: Summary of the 8 articles presenting resource use or cost data on multiple aspects for randomised clinical trials.

Author and year	Publication content in terms of empirical trial costs and resource use	Primary results regarding resource and cost data (based on abstract; in USD 2017)
Sertkaya et al., 2016 [25]	Aggregate cost data of different phases (I-III) from three proprietary data bases (Medidata solutions). Cost per RCT across therapeutic areas with aggregated cost details (per patient and per site) from the pharmaceutical industry (2004-2012). No resource data available.	Phase 2: USD 9.1 Mio (cardiovascular) – 25.4 Mio (hematology). Phase 3: USD 14.9 Mio (dermatology) – 68.5 Mio (pain and anesthesia)
Getz et al., 2015 [18]	Tufts Center for the Study of Drug Development study among working group of 15 pharmaceutical companies, looked between 2011 and 2012 at study reports and analysis plans, identified and assessed costs of “core” and "noncore" procedures per phase II/III protocol. No resource data available.	Phase 2: USD 300,000 (13.1%) is spent in direct costs of noncore procedures Phase 3: USD 1.8 Mio (18.5%) is spent in direct costs of noncore procedures
Raftery et al., 2015 NHS HTA Report [19]	Questionnaire used to assess planned and actual costs of RCTs funded by the NHS HTA programme (1995-2005). Provides mean cost per patient per year and other cost components (e.g. costs of “statistician input). “Research cost, as funded by the HTA programme, accounted for some 70% of the total cost…” No resource data available.	No abstract. Mean annual cost per patient: USD 3,062 – 9561
Zhu et al., 2013 [20]	A large trial (n=154,901) assessing different screening strategies for prostate, lung, colorectal and ovarian cancer. The trial was funded by the National Cancer Institute Division of Cancer Prevention. Besides overall costs, costs are separately listed for screening centers, coordinating center and collection of biological samples. No resource data are listed.	The overall costs over 20 years were USD 494 Mio with approximately USD 40.3 Mio for the collection, processing and storage of biospecimens.
Roche et al., 2002 [21]	In 1996 Clinical Research Associates (n=83) collected timing observations of 41 cancer-trial tasks (156 subtasks) over 30 consecutive days. Time efforts for all trial stages were obtained. No cost data available.	“Industry-sponsored studies had significantly higher overall mean times than did local and cooperative group studies. Early-phase studies required more time than did phase III trials.”
Evans et al., 2000 [22]	Retrospective estimate of resource use and cost of two phase II trials in lung cancer. Only site specific costs available in 1993 Canadian dollars. Resource and cost data for the project development phase (e.g. development of study protocol) and the phase “after last patient out“ not available.	Taxotere trial: USD 1,755 per patient Gemcitabine trial: USD 2,973 per patient
Thornquist et al., 1993 [23]	Costs and resource use of a large academic prevention trial (i.e. of β-carotene and vitamin A in decreasing the incidence of lung cancer in two populations at high risk) conducted between 1988 and 1991. Resource use (i.e. time effort) not presented in results.	The total direct costs over 13 years was approximately USD 80 Mio
Eri et al., 1992 [24]	Cost and extra time needed for patient recruitment and on site trial tasks in one academic hospital conducted between 1989 and 1991.	The costs for the hospital to carry out a trial with 65 evaluable is approximately USD 127,000

Abbreviations: RCT= Randomised clinical trial; NHS= National Health Service; HTA= Health Technology Assessment

Articles presenting resource or cost data for one specific aspect of RCTs

A total of 30 articles [26-55] described the costs of recruitment for an RCT (Table 3). Print advertisement (described in 14 articles), mail-based invitation (11 articles), and radio advertisement (6 articles) were the most commonly assessed recruitment methods. Phone screening was the cheapest recruitment method (USD 116, range USD 98-1,146 per recruited patient) and combining different media the most expensive (USD 2,058; range USD 91-16,399 per recruited patient). Fifteen articles reported the overall costs for recruitment independent of any recruitment strategy. The median costs for recruiting a patient were USD 409 (range: USD 41-6,990).

Table 3: Summary of the 30 articles presenting costs to recruit one patient into a randomised controlled trial according to different recruitment strategies.

Recruitment strategy	Number of studies (total n=30)	Median number of recruited patients (range)	Median costs per recruited patient in USD 2017 (range)	References
Print advertisement	14	19 (4-141)	580 (43-2,254)	[27, 29, 30, 35, 37-41, 43-46, 50]*
Mail invitation	11	82 (4-337)	228 (15-1,116)	[28, 32, 36, 38, 40, 41, 43-45, 53, 55]*
Radio advertisement	6	5 (3-38)	627 (204-5,243)	[27, 29, 35, 38, 43, 50]*
General practitioner referral	5	18 (4-107)	541 (9-2716)	[32, 37, 40, 50, 55]
Billboard/posters	5	15 (2-36)	1134 (64-12,538)	[27, 37, 45, 50, 55]
Community workers (Church, social service)	5	5 (1-152)	642 (22-1938)	[35, 37, 40, 41, 44]
Internet advertisement	4	11 (2-37)	198 (155-2,621)	[27, 29, 40, 43]*
Phone screening	4	192 (87-347)	116 (98-1146)	[36, 51-53]
Public transport advertisement	3	2 (2-26)	1,295 (1,252-7,750)	[27, 29, 32]
Combination of different media (i.e. radio, newspaper, posters)	3	27 (1-75)	2058 (91-16,399)	[31, 32, 55]
TV advertisement	2	24 (5-174)	440 (69-3,548)	[43, 56]*
Manual chart screening and algorithm search	1**	16 (11-20)	297 (227-366)	[34]
Overall costs of recruitment	15	164 (15-4249)	409 (41-6,990)	[26, 30, 33, 38, 40, 42-50, 54]*

* One article [43] reports on two separate datasets from two RCTs.
**One article with the two mentioned recruitment strategies (i.e. manual chart screening and algorithm search)

Five articles assessed other specific cost aspects of RCTs. Two articles evaluated the costs of protocol procedures including protocol amendments [57, 58], one article estimated the costs of adverse event procedures in an HIV trial [59], one calculated the average costs of a hospital pharmacy among all trials where the pharmacy has been involved [60], and one article assessed the costs of "GCP-related activities" in an RCT [61]. A comprehensive overview of these studies is listed in Appendix Table S1.

Articles presenting overall costs

Sixteen articles presented overall costs for one or several RCTs [20, 23, 25, 62-74] (Table 4). The methods to estimate these cost data were not reported at all in 7 articles (38%), or little detailed and intangible in 5 articles (31%; e.g. personal communication; estimated based on final study reports). Three out of these 16 articles were published editorials or comments. The costs of the cheapest RCT was listed by Detsky [73] and was a trial published in 1983 assessing the efficacy of propranolol in 24 patients with esophagael varices (overall costs USD 209,202) [75]. The most expensive trial was a cancer screening trial enrolling 154,901 patients over a period of eight years and conducted a follow-up of 13 years (overall costs USD 494 Mio) [20]. The data from this trial were used and presented in a total of 335 publications. The lowest costs per patient (USD 43) were reported in 2014 for a trial using routinely collected data as registry embedded trial (TASTE trial) that assessed thrombus aspiration during myocardial infarction in 7244 patients [66]. Endpoints were completely assessed trough national registries. The highest costs per patient (USD 103,254) were reported for a lipid research trial conducted in 1983 [73]. The trial randomised 3,806 men with primary hypercholesterolemia to bile acid sequestrant cholestyramine resin or placebo. Median follow-up was 7.4 years and participants visited the clinics every two months.

Table 4: Summary of the 16 articles presenting overall costs for randomised controlled trials

Reference	Description/medical field	Year RCTs conducted	Sample size	Overall costs in USD 2017 (published costs)	Costs per patient in USD 2017 (published costs per patient in USD if not indicated differently)	Methods of cost assessment
Larson et al., 2016 [63]	Three multicentered, multinational HIV trials primarily funded by the National Institute of Allergy and Infectious Diseases	2002-2016	4,150 - 5,472	101.7 Mio-151.6 Mio (71.9 Mio – 111.9 Mio	24,512-27,702 (17,313-20,441)	NR
Sertkaya et al., 2016 [25]	Aggregated data from proprietary databases (Medidata Solutions) for pharmaceutical clinical trial costs were analysed and mean costs for RCT costs of different medical fields calculated.	2004-2012	NR	Phase II: 9.0 Mio – 25.4 Mio (7.0 Mio-19.6 Mio) Phase III: 14.9 Mio – 68.5 Mio (11.5 Mio – 52.9 Mio)	NA	Aggregate cost data from three proprietary data bases (Medidata solutions) for pharmaceutical industry trials were analysed.
Anguera et al., 2015 [62]	Fully mobile automated RCT randomising depressive patients to two different mobile phone applications	2014-2015	1,098	323,624 (314,264)	295 (286)	Cost estimated using a "total study cost approach" factoring in β testing, staff time and efforts beyond those payments made for recruitment and participant remuneration.
Roth et al., 2014 [65]	NIH funded women's health initiative estrogen plus progestin clinical trial conducted in 40 centres in the United States	1997-2002; (costs in USD 2012)	16,608	276.1 Mio (260 Mio)	16,624 (15,655)	NR
Chit et al., 2013 [67]	Data from 24 influenza vaccine developers and a total of 39 vaccines. An average trial was modelled.	2000-2012	Phase II: 517 Phase III: 4,461	Phase II: 2.3 Mio (CAD 2.3 Mio) Phase III: 15.1 Mio (CAD 15.4 Mio)	Phase II: 4,354 (CAD 4,505) Phase III: 3,379 (CAD 3,457)	Cost estimates from influenza research group at the Canadian Center for Vaccinology.
Zhu et al., 2013 [20]	RCT assessing different screening strategies for prostate, lung, colorectal and ovarian cancer.	1993-2001 (costs in USD 2011)	154,901	494 Mio (454 Mio)	3,189 (2,931)	NR
Light et al., 2009 [69]	Estimates of drug development costs including trials for different	1987-2003; (costs in	Phase II: 184-	Phase II: 1.5 Mio-3.3 Mio (1.3 Mio –	Phase II: NA Phase III: 2,651-	Cost estimates based on searches in U.S. Patent and

	phases of two rotavirus vaccines (from Merck and GSK, respectively).	USD 2008)	2,464 Phase III: 63,225 – 68,083	2.9 Mio) Phase III: 167.6 Mio-280.9 Mio (148.0 Mio – 248.1 Mio)	4,129 (2,340 – 3,646)	Trademark Office, the U.S. SEC EDGAR database, Medline, periodicals, and corporate websites and interviews with senior researchers.
Baker-Smith et al., 2008 [70]	Evaluations of Trials from nine oral pediatric antihypertensive drugs submitted to the FDA	NR (submitted to FDA: 1997 – 2004)	110 - 441	1.8 Mio-17.5* Mio (1.3 Mio – 12.9* Mio)	10,475-49,412 (7,605 – 37,291)	Cost estimates generated by using commercially available trial cost estimator software to generate approximations of the cost of performing a clinical trial.
Li et al., 2007 [71]	Sixteen trials from the Pediatric Exclusivity program from different medical fields.	NR (costs in USD 2005)	25 - 404	2.1 Mio-16.2 Mio (1.7 Mio – 12.9* Mio)	17,532-94,482 (14,044 – 75,686)	Costs estimates based on final study reports.
Thornquist et al., 1993 [23]	Prevention trial assessing if β-carotene and vitamin A decreasing the incidence of lung cancer in two populations at high risk	1988-1991 (costs in USD 1989)	17,000	80.7 Mio (40.9 Mio)	4,747 (2,406)	Cost estimates were based on surveying staff about average time per task and materials required. A developed spreadsheet was used to account for all tasks and possible costs.
Drummond et al., 1992 [72]	Diabetic retinopathy RCT	1972-1975	1732	61.2 Mio (10.5 Mio)	35,359 (6,062)	NR
Detsky 1989 [73]	Cost effectiveness from 7 RCT (5 cardiovascular trials; 2 nutritional/metabolic trials).	1981-1983 (Published: 1982-1986)	24 – 6,400	209,202-367.2Mio (78,000 – 150 Mio)	4,359-103,254 (1,625 – 42,182))	Cost estimates based on publication (2 trials), funding source (2 trials), expert opinion (1 trial), and the author's assessment (2 trials).
Detsky 1985 [74]	Two secondary infarction prevention trials. 1. RCT: Anturane Re-infarction trial, conducted in Canada and the United States; 2. RCT: Aspirin Myocardial Infarction Study, 30 centers in the United States.	1975-1979	1,629 - 4,524	22.7 Mio- 77.0 Mio (5 Mio – 17 Mio)	13,909-17,029 (3,069 - 3,758)	Cost estimates based on "personal communication"
Editorials/Comments:						
Muennig, 2015	RCT about "patient activation" to	NR (costs	NR	283,207	-	NR

(Commentary) [64]	confront their providers when they ordered screening for mammography or prostate-cancer.	in USD 2014)		(275,000)		
Lauer et al., 2014 (Editorial) [66]	"Event-driven" trial assessing thrombus aspiration during ST-elevation myocardial infarction (TASTE) conducted in 31 centres in Sweden, Finland and Denmark.	NR	7,244	313,964 (300,000)	43 (41)	NR
Callaway, 2012 (*NATURE*-News) [68]	NIH funded trial to assess whether chelation therapy (TACT) reduces cardiovascular events. Conducted at 134 US and Canadian sites.	2003-2011	1,708	41.9 Mio (31.6 Mio)	24,514 (18,501)	NR

* Trial costs of one RCT are presented both in the publication by Baker-Smith and colleagues [70] and the publication by Li et al [71].
Abbreviations: RCT= randomised clinical trial; NIH=National Institutes of Health; FDA=US Food and Drug Administration; NR=not reported; NA=not assessable

Discussion

Our systematic literature search of over 6000 titles and abstracts identified 56 articles that presented some cost or resource use data of RCTs in any form but we found no article assessing resource use and associated cost data for all aspects of an RCT (including detailed empirical resource use during the planning, conduct, and reporting of an RCT, and the associated costs).

A large proportion of the included articles assessed the costs to recruit one participant into an RCT. Presumably this reflects the fact that recruitment of a sufficiently large number of patients is considered one of the biggest challenges by many investigators [15, 76], but also the fact that evaluating the costs of patient recruitment is probably a relatively simple study-within-a-trial (SWAT; 29 out of 30 articles). Our pooled results from 15 articles indicate a large range of recruitment-costs (USD 41 – 6,990) which reflects the large diversity of RCTs recruiting various patient populations (e.g. patients with chronic depression; diabetic patients; rainwater drinkers) and allocating them to a large variety of interventions. This diversity can also be seen in the RCTs that presented overall costs. Due to this high diversity of medical fields, interventions, necessary measurements and follow-up times it is difficult to generalise trial costs. One observation in our study was that some of the oldest RCTs were very costly (i.e. USD 61 Mio [72], USD 77 Mio [74], and USD 367 Mio [73]; Table 4), which is in contrast to the current opinion that RCT cost are on constant rise.

To the best of our knowledge this is the first systematic review addressing the existing evidence on RCT resource use and associated costs. Although we consider our search approach comprehensive (3 electronic databases: MEDLINE, EMBASE, and HealthSTAR) we might have overlooked articles in journals not indexed in these electronic search databases or sporadic cost information in main texts, with no related hint in the title, abstract, or keywords. We expect the most impactful journals and articles to be covered by the databases searched. Further, we acknowledge that our survey mostly portrays costs of RCTs in high income nations that may not necessarily overlap with those of low- or middle-income countries. We tried to make the available data comparable by translating all cost data into USD 2017. We had originally planned to use purchasing power parities to translate costs of studies that were conducted outside of the United States, into USD. However, due to the fact that many studies were conducted in several countries and some of the RCTs did not state where they were conducted, this would have introduced a relevant amount of inconsistency, and was therefore omitted.

To date, no similar systematic review exists to which we could compare our results. A short report published in 2017 by Martin et al. [77] identified similar gaps in the evidence base of RCTs. The authors undertook an analysis of cost data from 726 pharmaceutical interventional trials, resulting in a median cost of USD 8.6 million for phase II trials and USD 21.4 million for phase III trials. However, again, the report lacks a transparent methodology and presents overall cost numbers only.

Based on our systematic review, cost data for industry funded RCTs are sparse and little detailed. For RCTs funded by academic sponsors there seems to be no comprehensive analysis of the costs available to date. Therefore, we composed a detailed list of costing items applicable to RCTs, collected empirical data on resource use and associated costs for two RCTs conducted within our network of collaborators, and summarised the results in a case report *(Speich et al., submitted together with this manuscript to the JCE)*. With this we meant to set an example for a transparent and comprehensive reporting of actual resource use and costs of RCTs. Transparently published resource use and costs may support researchers and funders alike to prospectively plan and monitor the costs associated with the conduct of RCTs in the future.

Conclusion

While the present systematic review revealed several studies on costs or resource use for specific aspects of RCTs or articles providing estimates for total costs of individual RCTs, we could not identify any article reporting detailed empirical data on resource use and costs for all aspects of a RCT. Accessible empirical resource use and cost data are urgently needed to optimise the cost-effectiveness of RCTs and ultimately make the current research environment more sustainable.

References

[1] Morgan S, Grootendorst P, Lexchin J, Cunningham C, Greyson D. The cost of drug development: a systematic review. Health Policy. 2011;100:4-17.
[2] Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, Lindborg SR, et al. How to improve R&D productivity: the pharmaceutical industry's grand challenge. Nat Rev Drug Discov. 2010;9:203-14.
[3] DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: New estimates of R&D costs. J Health Econ. 2016;47:20-33.
[4] Light DW, Warburton R. Demythologizing the high costs of pharmaceutical research. The London School of Economics and Political Science. 2011; 1745-8552 BioSocieties 1–17. www.palgrave-journals.com/biosoc/ (accessed August 25, 2016).
[5] Adams CP, Brantner VV. Estimating the cost of new drug development: is it really 802 million dollars? Health Aff (Millwood). 2006;25:420-8.
[6] Zimmer AS, Steeg PS. Meaningful prevention of breast cancer metastasis: candidate therapeutics, preclinical validation, and clinical trial concerns. J Mol Med. 2015;93:13-29.
[7] Canada Clinical Trial Summit. Canadian Institute for Health Research. 2011.
[8] Duley L, Antman K, Arena J, Avezum A, Blumenthal M, Bosch J, et al. Specific barriers to the conduct of randomized trials. Clin Trials. 2008;5:40-8.
[9] Collins R, MacMahon S. Reliable assessment of the effects of treatment on mortality and major morbidity, I: clinical trials. Lancet. 2001;357:373-80.
[10] Collier R. Rapidly rising clinical trial costs worry researchers. Cmaj. 2009;180:277-8.
[11] Pei H, Wu Y, Wei Y, Yang Y, Teng S, Zhang H. Remote ischemic preconditioning reduces perioperative cardiac and renal events in patients undergoing elective coronary intervention: a meta-analysis of 11 randomized trials. PLoS ONE. 2014;9:e115500.
[12] Sinha G. European move affects academic trials research. J Natl Cancer Inst. 2006;98:1100-1.
[13] Eisenstein EL, Collins R, Cracknell BS, Podesta O, Reid ED, Sandercock P, et al. Sensible approaches for reducing clinical trial costs. Clin Trials. 2008;5:75-84.
[14] Bosch X. Europe's restrictive rules strangling clinical research. Nat Med. 2005;11:1260-.
[15] Kasenda B, von Elm E, You J, Blumle A, Tomonaga Y, Saccilotto R, et al. Prevalence, characteristics, and publication of discontinued randomized trials. JAMA. 2014;311:1045-51.
[16] Bloomberg. https://www.bloomberg.com/markets/currencies; [Accessed 27 July 2017].
[17] Mirghani H, Amen F, Blanchard P, Moreau F, Guigay J, Hartl DM, et al. Treatment de-escalation in HPV-positive oropharyngeal carcinoma: Ongoing trials, critical issues and perspectives. Int J Cancer. 2015;136:1494-503.
[18] Getz KA, Stergiopoulos S, Marlborough M, Whitehill J, Curran M, Kaitin KI. Quantifying the magnitude and cost of collecting extraneous protocol data. Am J Ther. 2015;22:117-24.
[19] Raftery J, Young A, Stanton L, Milne R, Cook A, Turner D, et al. Clinical trial metadata: defining and extracting metadata on the design, conduct, results and costs of 125 randomised clinical trials funded by the National Institute for Health Research Health Technology Assessment programme. Health Technol Assess. 2015;19:1-138.
[20] Zhu CS, Pinsky PF, Kramer BS, Prorok PC, Purdue MP, Berg CD, et al. The prostate, lung, colorectal, and ovarian cancer screening trial and its associated research resource. J Natl Cancer Inst. 2013;105:1684-93.
[21] Roche K, Paul N, Smuck B, Whitehead M, Zee B, Pater J, et al. Factors affecting workload of cancer clinical trials: results of a multicenter study of the National Cancer Institute of Canada Clinical Trials Group. J Clin Oncol. 2002;20:545-56.
[22] Evans WK, Dahrouge S, Stapleton J, Quinn C, Pollock D, Waterfield B, et al. An estimate of the cost of conducting phase II trials in lung cancer. Lung Cancer. 2000;28:85-95.
[23] Thornquist MD, Urban N, Tseng A, Edelstein C, Lund B, Omenn GS. Research cost analyses to aid in decision making in the conduct of a large prevention trial, CARET. Carotene and Retinol Efficacy Trial. Control Clin Trials. 1993;14:325-39.
[24] Eri LM, Tveter KJ. Patient recruitment to and cost of a prospective trial of medical treatment for benign prostatic hyperplasia. Eur Urol. 1992;22:9-13.

[25] Sertkaya A, Wong HH, Jessup A, Beleche T. Key cost drivers of pharmaceutical clinical trials in the United States. Clin. 2016;13:117-26.
[26] Kira A, Glover M, Walker N, Bauld L. Recruiting pregnant indigenous women who smoke into a high contact incentivized cessation trial: A feasibility study. Nicotine and Tobacco Research. 2016;18:2036-40.
[27] Tompkins DA, Sides JA, Harrison JA, Strain EC. Recruitment techniques for alcohol pharmacotherapy clinical trials: A cost-benefit analysis. Addict Disord Their Treat. 2015;14:211-9.
[28] Gleason K, Shin D, Rueschman M, Weinstock T, Wang R, Ware JH, et al. Challenges in recruitment to a randomized controlled study of cardiovascular disease reduction in sleep apnea: an analysis of alternative strategies. Sleep. 2014;37:2035-8.
[29] Krusche A, Rudolf von Rohr I, Muse K, Duggan D, Crane C, Williams JMG. An evaluation of the effectiveness of recruitment methods: the staying well after depression randomized controlled trial. Clin. 2014;11:141-9.
[30] Hapca A, Jennings CG, Wei L, Wilson A, MacDonald TM, Mackenzie IS. Effectiveness of newspaper advertising for patient recruitment into a clinical trial. Br J Clin Pharmacol. 2014;77:1064-72.
[31] Redmond NM, Hollinghurst S, Costelloe C, Montgomery AA, Fletcher M, Peters TJ, et al. An evaluation of the impact and costs of three strategies used to recruit acutely unwell young children to a randomised controlled trial in primary care. Clin. 2013;10:593-603.
[32] Dew A, Khan S, Babinski C, Michel N, Heffernan M, Stephan S, et al. Recruitment strategy cost and impact on minority accrual to a breast cancer prevention trial. Clin. 2013;10:292-9.
[33] Penberthy LT, Dahman BA, Petkov VI, DeShazo JP. Effort required in eligibility screening for clinical trials. J Oncol Pract. 2012;8:365-70.
[34] Beauharnais CC, Larkin ME, Zai AH, Boykin EC, Luttrell J, Wexler DJ. Efficacy and cost-effectiveness of an automated screening algorithm in an inpatient clinical trial. Clin. 2012;9:198-203.
[35] Graham AL, Lopez-Class M, Mueller NT, Mota G, Mandelblatt J. Efficiency and cost-effectiveness of recruitment methods for male Latino smokers. Health Educ Behav. 2011;38:293-300.
[36] Beaton SJ, Sperl-Hillen JM, Worley AV, Fernandes OD, Baumer D, Hanson AM, et al. A comparative analysis of recruitment methods used in a randomized trial of diabetes education interventions. Contemp Clin Trials. 2010;31:549-57.
[37] Butt DA, Lock M, Harvey BJ. Effective and cost-effective clinical trial recruitment strategies for postmenopausal women in a community-based, primary care setting. Contemp Clin Trials. 2010;31:447-56.
[38] Cambron JA, Dexheimer JM, Chang M, Cramer GD. Recruitment methods and costs for a randomized, placebo-controlled trial of chiropractic care for lumbar spinal stenosis: a single-site pilot study. J Manipulative Physiol Ther. 2010;33:56-61.
[39] Grov C, Bux D, Jr., Parsons JT, Morgenstern J. Recruiting hard-to-reach drug-using men who have sex with men into an intervention study: lessons learned and implications for applied research. Subst Use Misuse. 2009;44:1855-71.
[40] Raynor HA, Osterholt KM, Hart CN, Jelalian E, Vivier P, Wing RR. Evaluation of active and passive recruitment methods used in randomized controlled trials targeting pediatric obesity. Int J Pediatr Obes. 2009;4:224-32.
[41] Rodrigo S, Sinclair M, Cunliffe D, Leder K. Effectiveness and cost of recruitment strategies for a community-based randomised controlled trial among rainwater drinkers. BMC Med Res Methodol. 2009;9:51.
[42] Walton JN, MacEntee MI. Screening and enrolling subjects in a randomized clinical trial involving implant dentures. Int J Prosthodont. 2008;21:210-4.
[43] Galbreath AD, Smith B, Wood P, Forkner E, Peters JI. Cumulative recruitment experience in two large single-center randomized, controlled clinical trials. Contemp Clin Trials. 2008;29:335-42.
[44] Sadler GR, Ko CM, Malcarne VL, Banthia R, Gutierrez I, Varni JW. Costs of recruiting couples to a clinical trial. Contemp Clin Trials. 2007;28:423-32.

[45] Robinson JL, Fuerch JH, Winiewicz DD, Salvy SJ, Roemmich JN, Epstein LH. Cost effectiveness of recruitment methods in an obesity prevention trial for young children. Prev Med. 2007;44:499-503.
[46] Perri R, Wollin S, Drolet N, Mai S, Awad M, Feine J. Monitoring recruitment success and cost in a randomized clinical trial. Eur J Prosthodont Restor Dent. 2006;14:126-30.
[47] Gismondi PM, Hamer DH, Leka LS, Dallal G, Fiatarone Singh MA, Meydani SN. Strategies, time, and costs associated with the recruitment and enrollment of nursing home residents for a micronutrient supplementation clinical trial. J Gerontol A Biol Sci Med Sci. 2005;60:1469-74.
[48] Rubin RR, Fujimoto WY, Marrero DG, Brenneman T, Charleston JB, Edelstein SL, et al. The Diabetes Prevention Program: recruitment methods and results. Control Clin Trials. 2002;23:157-71.
[49] MacEntee MI, Wyatt C, Kiyak HA, Hujoel PP, Persson RE, Persson GR, et al. Response to direct and indirect recruitment for a randomised dental clinical trial in a multicultural population of elders. Community Dent Oral Epidemiol. 2002;30:377-81.
[50] Kaakko T, Murtomaa H, Milgrom P, Getz T, Ramsay DS, Coldwell SE. Recruiting phobic research subjects: effectiveness and cost. Anesth Prog. 2001;48:3-8.
[51] Gill TM, McGloin JM, Gahbauer EA, Shepard DM, Bianco LM. Two recruitment strategies for a clinical trial of physically frail community-living older persons. J Am Geriatr Soc. 2001;49:1039-45.
[52] Miller NL, Markowitz JC, Kocsis JH, Leon AC, Brisco ST, Garno JL. Cost effectiveness of screening for clinical trials by research assistants versus senior investigators. J Psychiatr Res. 1999;33:81-5.
[53] Margitic S, Sevick MA, Miller M, Albright C, Banton J, Callahan K, et al. Challenges faced in recruiting patients from primary care practices into a physical activity intervention trial. Activity Counseling Trial Research Group. Prev Med. 1999;29:277-86.
[54] Bielski RJ, Lydiard RB. Therapeutic trial participants: where do we find them and what does it cost? Psychopharmacol Bull. 1997;33:75-8.
[55] Bjornson-Benson WM, Stibolt TB, Manske KA, Zavela KJ, Youtsey DJ, Buist AS. Monitoring recruitment effectiveness and cost in a clinical trial. Control Clin Trials. 1993;14:52S-67S.
[56] Tompkins DA, Harrison JA, Strain EC. What is the most cost-effective advertising strategy for alcohol pharmacotherapy clinical trials? Drug and Alcohol Dependence. 2015;146:e5.
[57] Getz KA, Stergiopoulos S, Short M, Surgeon L, Krauss R, Pretorius S, et al. The impact of protocol amendments on clinical trial performance and cost. Therapeutic Innovation and Regulatory Science. 2016;50:436-41.
[58] Getz K. Improving protocol design feasibility to drive drug development economics and performance. Int J Environ Res Public Health. 2014;11:5069-80.
[59] Chou VB, Omer SB, Hussain H, Mugasha C, Musisi M, Mmiro F, et al. The costs associated with adverse event procedures for an international HIV clinical trial determined by activity-based costing. J Acquir Immune Defic Syndr. 2007;46:426-32.
[60] Idoate A, Ortega A, Carrera FJ, Aldaz A, Giraldez J. Cost-evaluation model for clinical trials in a hospital pharmacy service. Pharm World Sci. 1995;17:172-6.
[61] Funning S, Grahnen A, Eriksson K, Kettis-Linblad A. Quality assurance within the scope of Good Clinical Practice (GCP) - What is the cost of GCP-related activities? A survey within the Swedish Association of the Pharmaceutical Industry (LIF)'s members. Quality Assurance Journal. 2009;12:3-7.
[62] Anguera JA, Jordan JT, Castaneda D, Gazzaley A, Arean PA. Conducting a fully mobile and randomised clinical trial for depression: access, engagement and expense. BMJ Innovations. 2016;2:14-21.
[63] Larson GS, Carey C, Grarup J, Hudson F, Sachi K, Vjecha MJ, et al. Lessons learned: Infrastructure development and financial management for large, publicly funded, international trials. Clin. 2016;13:127-36.
[64] Muennig PA. How Automation Can Help Alleviate the Budget Crunch in Public Health Research. Am J Public Health. 2015;105:e19-22.

[65] Roth JA, Etzioni R, Waters TM, Pettinger M, Rossouw JE, Anderson GL, et al. Economic return from the Women's Health Initiative estrogen plus progestin clinical trial: a modeling study. Ann Intern Med. 2014;160:594-602.
[66] Lauer MS, Bonds D. Eliminating the "expensive" adjective for clinical trials. Am Heart J. 2014;167:419-20.
[67] Chit A, Parker J, Halperin SA, Papadimitropoulos M, Krahn M, Grootendorst P. Toward more specific and transparent research and development costs: the case of seasonal influenza vaccines. Vaccine. 2014;32:3336-40.
[68] Callaway E. Chelation-therapy heart trial draws fire. Nature. 2012;491:313-5.
[69] Light DW, Andrus JK, Warburton RN. Estimated research and development costs of rotavirus vaccines. Vaccine. 2009;27:6627-33.
[70] Baker-Smith CM, Benjamin DK, Jr., Grabowski HG, Reid ED, Mangum B, Goldsmith JV, et al. The economic returns of pediatric clinical trials of antihypertensive drugs. Am Heart J. 2008;156:682-8.
[71] Li JS, Eisenstein EL, Grabowski HG, Reid ED, Mangum B, Schulman KA, et al. Economic return of clinical trials performed under the pediatric exclusivity program. Jama. 2007;297:480-8.
[72] Drummond MF, Davies LM, Ferris FL, 3rd. Assessing the costs and benefits of medical research: the diabetic retinopathy study. Soc Sci Med. 1992;34:973-81.
[73] Detsky AS. Are clinical trials a cost-effective investment? Jama. 1989;262:1795-800.
[74] Detsky AS. Using economic analysis to determine the resource consequences of choices made in planning clinical trials. J Chronic Dis. 1985;38:753-65.
[75] Burroughs AK, Jenkins WJ, Sherlock S, Dunk A, Walt RP, Osuafor TO, et al. Controlled trial of propranolol for the prevention of recurrent variceal hemorrhage in patients with cirrhosis. N Engl J Med. 1983;309:1539-42.
[76] Sung NS, Crowley WF, Jr., Genel M, Salber P, Sandy L, Sherwood LM, et al. Central challenges facing the national clinical research enterprise. JAMA. 2003;289:1278-87.
[77] Martin L, Hutchens M, Hawkins C, Radnov A. How much do clinical trials cost? Nat Rev Drug Discov. 2017;16:381-2.

Appendix

Search strategy

Medline

Database: OVID Medline Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R) 1946 to November 30 2016.

Search Strategy:
---
1 exp Clinical Trials as Topic/ec
2 exp Clinical Trials as Topic/
3 ((clinical or random*) adj2 trial*).mp.
4 2 or 3
5 "costs and effectiveness".ti,ab,kf.
6 ((cost or costs) adj3 research).mp.
7 financial management.mp. or Financial Management/
8 (clinical trial* cost* or controlled trial* cost* or trial cost*).mp.
9 opportunity cost*.mp.
10 (cost driver* or cost factor*).mp.
11 (research adj2 cost*).mp.
12 (cost* adj2 conduct*).mp.
13 or/5-12
14 13 and 4
15 1 or 14
16 exp "Costs and Cost Analysis"/
17 exp *"Costs and Cost Analysis"/ and 2
18 exp *Clinical Trials as Topic/ and 16
19 17 or 18
20 19 not 15
21 15 or 19
****

EMBASE

Database: Embase <1974 to November 30 2016>
Search Strategy:
---
1 exp Clinical Trials as Topic/ec
2 exp Clinical Trials as Topic/
3 ((clinical or random*) adj2 trial*).mp.
4 2 or 3
5 "costs and effectiveness".ti,ab,kf.
6 ((cost or costs) adj3 research).mp.
7 financial management.mp. or Financial Management/
8 (clinical trial* cost* or controlled trial* cost* or trial cost*).mp.
9 opportunity cost*.mp.
10 (cost driver* or cost factor*).mp.
11 (research adj2 cost*).mp.
12 (cost* adj2 conduct*).mp.
13 or/5-12
14 13 and 4
15 1 or 14
16 exp "Costs and Cost Analysis"/
17 exp *"Costs and Cost Analysis"/ and 2
18 exp *Clinical Trials as Topic/ and 16
19 17 or 18
20 19 not 15
21 15 or 19
****

HealthSTAR

Database: Ovid Healthstar <1966 to November 30 2016>
Search Strategy:
---

1 exp Clinical Trials as Topic/ec
2 exp Clinical Trials as Topic/
3 ((clinical or random*) adj2 trial*).mp.
4 2 or 3
5 "costs and effectiveness".ti,ab,kf.
6 ((cost or costs) adj3 research).mp.
7 financial management.mp. or Financial Management/
8 (clinical trial* cost* or controlled trial* cost* or trial cost*).mp.
9 opportunity cost*.mp.
10 (cost driver* or cost factor*).mp.
11 (research adj2 cost*).mp.
12 (cost* adj2 conduct*).mp.
13 or/5-12
14 13 and 4
15 1 or 14
16 exp "Costs and Cost Analysis"/
17 exp *"Costs and Cost Analysis"/ and 2
18 exp *Clinical Trials as Topic/ and 16
19 17 or 18
20 19 not 15
21 15 or 19

****

Table S1: Summary of the five articles assessing other specific resource use or cost aspects than recruitment to a randomized clinical trial

Reference	Year RCT conducted	Costing unit	Study description	Main results (cost data as published [not converted into USD 2017])
Idoate et al., 1995	1993-1994	Pharmacy services	Analysis of cost evaluation model to estimate cost of hospital pharmacy services, including fixed (common to all trials) and variable (peculiar to each trial) costs were determined for each step. Economic assessment of each item was based on cost of the materials and means used, cost of staff time, and cost of drug storage. The model was applied to 24 clinical trials at University Clinic of Navarra.	83% of all pharmacy costs were variable. Drug dispensing, stock management and return drugs account for 94% of the time expended. Approximate cost of pharmacy providing investigational services was USD 174 per patient.
Funning et al., 2009	Industry data from 2005	GCP-related activities	Electronic survey of ICH GCP-activities and their related costs in 47 of 60 member companies of the Swedish Association of the Pharmaceutical Industry.	In 97% (n=250) of phase III trials performed in Sweden in 2005, approximately 50% of the total budget was reported to be GCP-related, estimated at an average of USD 180 million. 50% of GCP-related cost was related to Source Data Verification, estimated at USD 90 million.
Chou et al., 2007	2005-2006	Adverse events procedures	Analysis of costs assosicated with adverse events procedures in a large HIV perinatal trial in Uganda by determining actual resource consumption using activity-based costing. Resources were organized into cost categories (e.g. personnel, patient care expenses, laboratory testing, equipment). Cost drivers were quantified and unit cost per adverse event was calculated.	In 18 months, there were 9028 adverse events with 970 reported as serious adverse events. Unit cost per adverse event was USD 101.97). Overall, adverse event-related costs represent 32% (USD 920,581 of USD 2,834,692) of all RCT expenses. Costs for personnel (USD 79.3) and patient care (USD 11.96) contributed the greatest proportion.
Getz 2014	Industry Data from 2002-2012	Study protocol procedures (core, required, standard)	Review of the results of two major Tuft Center for the Study of Drug Development studies quantifying the direct cost of conducting less essential and unnecessary protocol procedures in industry phase II-III studies and of implementing amendments to protocol designs. Data from 15-17 mid-sized and large pharmaceutical and biotechnology companies.	47.9% of total RCT budget on average spent on "core" procedures in phase II and III protocols, 22.7% of costs for "required" (regulatory compliance) and 12.0% for "standard" procedures. Completed protocols across all phases (I-IV) had an average of 2.3 amendments, requiring an average of 6.9 changes to the protocol. Total investigative site work burden to administer procedures supporting phase II, III, and IV protocols increased 73%, 56%, and 57% respectively during the ten year period of 2002 to 2012.
Getz et al., 2016	Industry data from 2010-	Protocol amendments	Analysis of data from 836 phase I-IV industry protocols to understand amendment prevalence , their impact, and	57% of protocols had at least one substantial amendment, and nearly half (45%) of these

	2013, follow up in 2015 of 2010 RCT		associated direct costs	amendments were deemed avoidable. Phase II and III protocols had a mean number of 2.2 and 2.3 global amendments, respectively. A total of 52 protocols were analysed to derive cost estimates of the direct costs to implement amendments. The median direct cost to implement a substantial amendment was USD 141,000) for a phase II protocol and USD 535,000 for a phase III protocol

Abbreviations: RCT=randomized clinical trial; ICH= International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use; GCP= Good clinical practice; HIV= human immunodeficiency virus

Manuscript IV: Cost and resource use evaluation of randomised clinical trials: a case study exemplifying standardised assessment using a comprehensive cost item list

Benjamin Speich[a,†], Belinda von Niederhäusern[b,†], Claudine Angela Blum[c,d], Jennifer Keiser[e], Nadine Schur[e] Thomas Fürst[e,g,h], Benjamin Kasenda[a,i], Mirjam Christ-Crain[c], Lars G. Hemkens[a], Christiane Pauli-Magnus[b], Matthias Schwenkglenks[f,j]*, Matthias Briel[a,k]* for the MAking Randomized Trials Affordable (MARTA) Group

[a]Basel Institute for Clinical Epidemiology and Biostatistics, Department of Clinical Research, University of Basel and University Hospital Basel, Switzerland
[b]Clinical Trial Unit, Department of Clinical Research, University of Basel and University Hospital Basel, Switzerland
[c]Endocrinology, Diabetology and Metabolism, Department of Internal Medicine and Department of Clinical Research, University Hospital Basel, Switzerland
[d]Medical University Clinic, Departments of General Internal and Emergency Medicine and Department of Endocrinology, Diabetology and Clinical Nutrition, Cantonal Hospital Aarau
[e]Department of Medical Parasitology and Infection Biology, Swiss Tropical and Public Health Institute, University of Basel, Basel, Switzerland
[f]Institute of Pharmaceutical Medicine, University of Basel, Basel, Switzerland
[g]Department of Epidemiology and Public Health, Swiss Tropical and Public Health Institute, Basel, Switzerland
[h]School of Public Health, Imperial College London, London, United Kingdom
[i]Department of Medical Oncology, University of Basel and University Hospital Basel, Switzerland
[j]Epidemiology, Biostatistics and Prevention Institute, University of Zürich, Zürich, Switzerland
[k]Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, ON, Canada

† These authors contributed equally to the work reported here and share first authorship.
* These authors contributed equally to the work reported here and share last authorship.

Status: Under review, Journal of Clinical Epidemiology

Abstract

Objectives: Randomised clinical trials (RCTs) are costly and published information on resource requirements for their conduct is limited. To identify suitable lever-points to make RCTs more cost-efficient and feasible, empirical data on resource use and associated costs are needed. We aimed to retrospectively assess resource use and detailed costs of two academic, investigator-initiated RCTs using a comprehensive list of cost items.

Study Design and Setting: The resource use of two investigator-initiated RCTs (Prednisone-Trial [NCT00973154] and Oxantel-Trial [ISRCTN54577342]) was empirically assessed in a standardized manner through semi-structured interviews and a systematically developed cost item list. Using information about yearly salaries, resource use was translated into costs. Additionally, we collected all "other costs" including fixed priced items. Overall costs as well as cost of different study phases were calculated.

Results: The personnel time used in the Prednisone-Trial trial was approximately 2,897 working days and the overall costs were calculated to be USD 2.3 million, which was USD 700,000 more than planned. In the Oxantel-Trial 798 working days were spent and the overall costs were as originally planned USD 100,000. Cost drivers were similar with recruitment delays explaining the additional costs in the Prednisone-Trial.

Conclusion: This case study provides an example of how to transparently assess resources and costs of RCTs, and presents detailed empirical data on type and magnitude of expenses. In the future, this model approach may serve others to plan, assess, or monitor resource use and costs of RCTs.

What is new?

Key Findings

- The two assessed randomised clinical trials (RCT) had a personnel time of 2,897 and 798 working days and total costs were USD 2.3 million and USD 100,000, respectively.
- Even though resource use and costs were substantially different, we identified the same cost drivers (salary costs, patient enrolment, treatment, and follow-up phase)

What this adds to what was known?

- To the best of our knowledge this is the first study using a comprehensive list of RCT cost items to provide detailed data on resource use and associated costs of academic, investigator-initiated RCTs
- This study exemplifies how resource use and cost data can be systematically collected and transparently presented to support others in the planning and monitoring of resources for RCTs.

What is the implication and what should change now?

- More studies of this type are needed to establish cost drivers of RCTs and identify lever-points for cost savings.

Introduction

The planning and conduct of randomised clinical trials (RCTs) are often complex and consume substantial resources [1-5]. Particularly in resource-limited settings such as academic clinical research the conduct of investigator-initiated RCTs is challenging [6]. To improve the feasibility and cost-effectiveness of RCTs, it is necessary to identify suitable lever-points for interventions based on empirical data on resource use and costs of relevant tasks in RCTs.

A systematic literature search about the resources and costs of RCTs identified 56 articles that presented some cost or resource use data of RCTs in any form (*Speich et al., submitted together with this manuscript to the JCE*). However, none of these articles reported on the detailed resource use and associated costs for all aspects of an entire RCT. Furthermore, while 16 articles contained numbers of overall costs of particular RCTs, the methodology of how these estimates were obtained remained unclear in 12 studies. There is consensus that more transparency in the methodology applied and the reporting of data on resource use (e.g. number of clinical procedures performed, hours administrative staff time used, and number of variables requiring source data verification) would be powerful to identify main drivers of the ever-rising costs of clinical research and development [7, 8].

Therefore, the aims of this study were to i) compile a comprehensive list of cost items applicable to RCTs and ii) retrospectively assess resource use and detailed costs of two academic, investigator-initiated RCTs [9, 10]. We further exemplify how costs and resource use can be systematically assessed for RCTs in general [7].

Methods

Comprehensive cost item list

We designed and conducted a systematic literature search of costing templates or cost item lists on MEDLINE and EMBASE via the Ovid interphase and EconLit on the 3rd November 2015. Our search was not restricted in terms of language, publication type, or publication date (Appendix A for detailed search strategy). In addition, we conducted a systematic search of costing templates on the internet (websites and linked information, see Appendix A for detailed search strategy). We extracted cost items associated with clinical trials and compiled them in an electronic, shareable database (googledocs.com). In addition to the systematic searches, we received two clinical trial budget templates from institutions involved in clinical research in Switzerland (one industry / one not-for-profit). All articles, websites, and

documents were searched for cost items following the principle of saturation, i.e. until no additional items emerged. For validation, we circulated the comprehensive item list among six experts from the pharmaceutical industry and five experts from academia. We iteratively discussed the suggestions and made adaptations to the item list where applicable. In the final list, we structured all cost items into direct and indirect costs. Lastly, we introduced functionalities (e.g. automatic calculation of staff rates multiplying with hours in costing sheets etc.) and developed a short instruction manual as well as a glossary for important template terminology.

Cost and resource use evaluation of two randomised clinical trials

Study selection

We selected two readily available RCTs for the following reasons: First, both were conducted within the close network of the authors, therefore relevant documents and information about costs (i.e. budgets, bills) were readily available. Second, both studies were published recently, meaning that the responsible persons could be interviewed about undocumented costs. Third, based on the different settings and the different pace of enrolment, empirical data from these two studies may give a hint about the cost range of investigator-initiated RCTs. Fourth, the two RCTs were of high methodological quality and published in high impact journals.

An overview of the study characteristics is given in Table 1. The first trial, conducted at seven centres in Switzerland from 2009 to 2015, randomised a total of 802 patients with community-acquired pneumonia to two different treatment arms ([NCT00973154]) [10]. For simplicity, we will refer to this trial as the “Prednisone-Trial” [10, 11]. Patients were randomly assigned to receive either adjunct prednisone 50 mg daily for seven days or placebo (all patients received antibiotics as standard care) [10]. The time from the beginning of the enrolment until the last patient finished the follow-up took 59 months. The initially planned time was 48 months.

The second RCT was planned and conducted within a collaboration of researchers from Switzerland and Tanzania and took place on Pemba Island, Tanzania, in 2012 (ISRCTN54577342) [9]. Children from two schools who were infected with intestinal worms (i.e. *Trichuris trichiura* and hookworm) were enrolled and randomly assigned to one of four treatment arms over a period of two months [9]. We will refer to this trial as the “Oxantel-Trial”.

Table 1: Characteristics of the two included randomised clinical trials for resource use and cost evaluation.

Study	**Prednisone-Trial [10]**	**Oxantel-Trial [9]**
Year study conducted	2009-2014	2012
Year study published	2015	2014
Setting	Seven hospitals across Switzerland	Two schools on Pemba Island, Tanzania
Sample size	802 participants	480 participants
Population and disease	Adults with community-acquired pneumonia	Children 6 to 14 years of age, infected with soil-transmitted helminths (i.e. *Trichuris trichiura* and hookworm)
Interventions	Prednisone Placebo	Oxantel pamoate and albendazole Oxantel pamoate Albendazole Mebendazole
Planned budget	USD 1.59 Mio.	USD 100,000
Planned time from start of enrolment until last patient out (in months)	48	2
Actual time from start of enrolment until last patient out (in months)	59	2
Primary outcome	Time to clinical stability (in days)	Cure rate and egg reduction rate

Resource and associated cost estimations

We conducted semi-structured interviews using our comprehensive cost item list (Appendix B) as a guideline to cover all relevant aspects of an RCT. We stratified all items by study phases, i.e. "*Project development and preparation*", "*Patient enrolment, treatment and follow-up*", "*After last patient out*", and "*Additional costs*". In all categories, we divided costs into (i) costs associated with human resources ("salary costs") based on the hours worked multiplied by the respective staff salary per hour; and (ii) "other costs" including fixed priced items (e.g.laboratory materials, approval fees) (Figure 1). As one presented RCT was conducted from 2009 until 2014 and the other in 2012 all costs were expressed as values of July 2012 USD (1 Swiss Franc = USD 1.055).

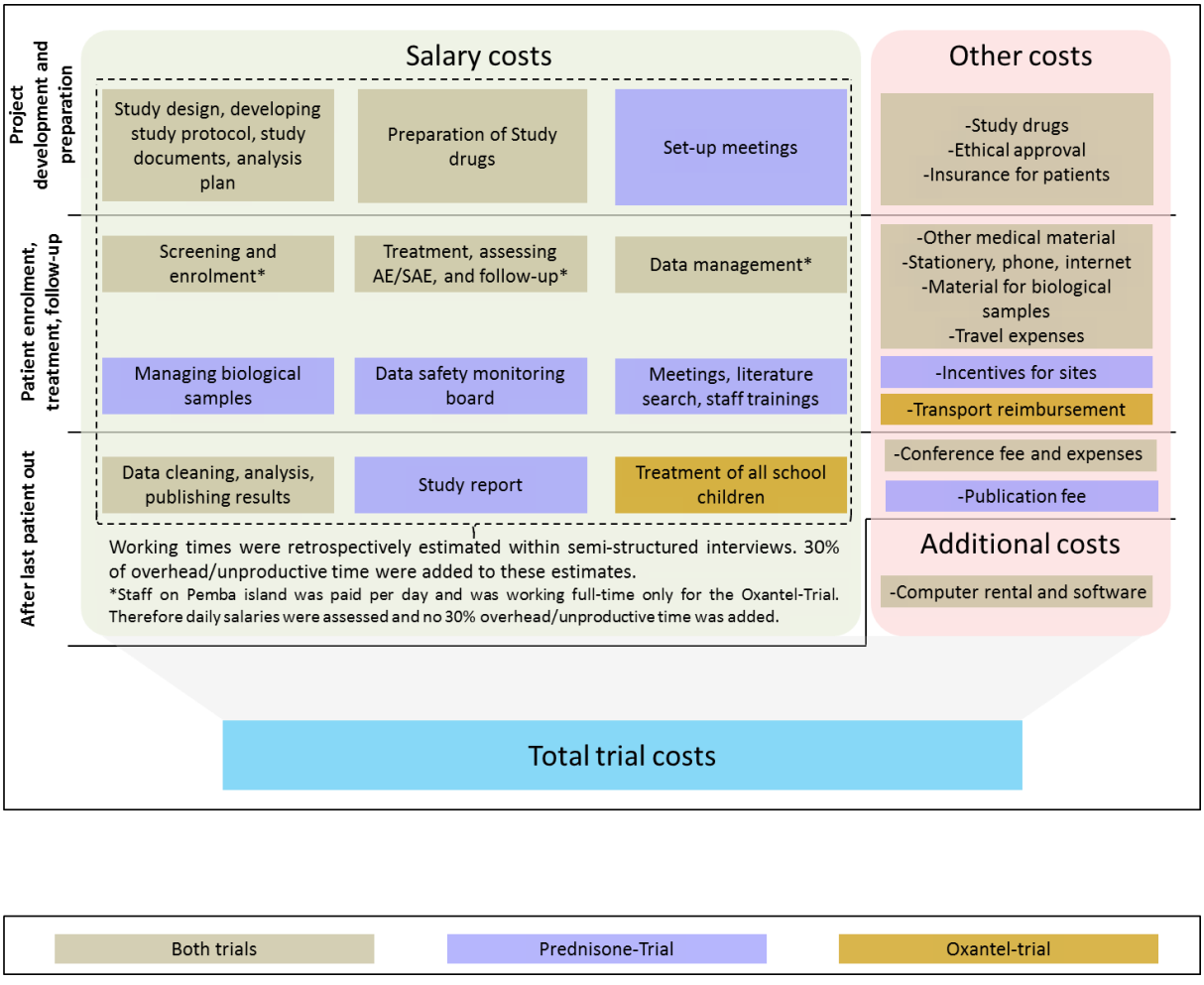


Figure 1: Categories of the cost and resource use evaluation, using the example of two randomised clinical trials.

Human resources

We retrospectively estimated human resource use based on available time log sheets and by conducting semi-structured interviews with study staff. Some of the authors were involved in one of the two RCTs (i.e. CAB, MCC, MB, CB, Prednisone-Trial; BS, JK, Oxantel-Trial) and were therefore well aware of the resource use associated with the specific stages of the RCTs.

Within the semi-structured interviews, we asked the involved persons what time effort (hours, working days) they had spent on each working step (e.g. preparation of specific documents, meetings, organising patient insurance, screening patient, enrolling patient, data management, statistical analysis, publishing results; see Appendix 1). Moreover, the interviewed person explained all their tasks within the study to assess if additional working time from other, non-pre-specified tasks had to be considered. To account for non-included time (e.g. walking between patients or buildings, discussing study) and for unproductive phases, we added 30% of overhead/unproductive time to each item [12, 13].

Of note, the human resources for *patient enrolment, treatment, and follow-up* could not be assessed in an identical way for the two RCTs. For the Prednisone-Trial it was not possible to identify all involved persons at each of the seven recruiting hospitals. We therefore extrapolated the data obtained from the main coordinating centre to the co-investigating centres in case no time log sheets were available. Within the Oxantel-Trial employees on Pemba Island, Tanzania were hired on a daily basis and were working full time only for the Oxantel-Trial. Therefore, daily salaries were used rather than costs based on human resource estimates. Consequently no 30% of overhead/unproductive time was added.

Human resources translated into salary costs

Salary tables from 2012 from the University Hospital Basel, from the Swiss Tropical and Public Health Institute (Swiss TPH), and the daily salaries for the Public Health Laboratory on Pemba Island were used to calculate total salary costs (including social benefits) based on estimates of human resource use. In detail, we used (i) daily salaries (Pemba Island, internal consulting at the Swiss TPH) or (ii) hourly salary rates, taking into account varying target hours per day depending on position of medical staff for all resource use in Switzerland. An overhead of 25% for the University Hospital, 20% for the Public Health Laboratory on Pemba Island, and 15% for the Swiss Tropical and Public Health Institute was applied. Resources and costs were then tabulated as working days and associated costs.

Other resources and costs

To evaluate other costs, we listed all expenditures which were explicitly made for the studied RCTs, including fees (e.g. ethical approval, trial registration, publication), expenses, study drugs, reimbursement for public transport, infrastructure (e.g. car, freezer), services (e.g. diagnosis from hospital laboratory) and materials (e.g. medical material, laboratory material, stationery). The vast majority of these costs could easily be assessed from bills or financial reports (e.g. to public funding agency). For some items the life expectancy had to be estimated (e.g. freezer). It was assessed how many days these materials and infrastructure were used and how much the items cost per day. More general, less trial specific infrastructure costs such as office space was included in the overhead costs of the salaries.

Analyses

We descriptively listed and categorised all resources used, their associated costs and proportional contribution to the total costs, as mentioned above. Main resource categories are listed in Figure 1; a more detailed list is presented in the Appendix (Appendix C-D).

To determine the robustness of our cost estimates and to assess to what extent the overall costs vary if single parameters are modified (different scenarios), a series of one-way sensitivity analyses was conducted [14]. The following scenarios were assessed: (i) No addition of 30% overhead/unproductive time to evaluate how costs decrease in case the added time was overestimated; (ii) salaries increase by 30%; (iii) increase and decrease in sample size of 50% (only influence on costs in *patient enrolment, treatment, follow-up,* study drugs, insurance, and data cleaning); (iv) increase and decrease of 50% for *project development and planning* costs; (v) 50% higher and lower costs *after last patient out*; (vi) increase and decrease of 50% for other costs as well as additional costs.

Results

Comprehensive cost item list

We identified eight relevant templates in our literature search and 17 templates in our internet search. All items were listed and we circulated the comprehensive item list amongst experts. Four out of six experts from industry and three out of five experts from academia provided feedback on the item list. The final item list included a total of 137 different cost items which were categorized into direct and indirect costs.

Direct study costs include those required to complete the work outlined in the protocol. These can be further subdivided into *fixed costs* related to the cost of supporting the study and *variable* costs related directly to the sample size of the study [15, 16]. We further stratified direct costs by research phase, i.e. i) planning and preparation, ii) conduct, and iii) after last patient out.

Results from the case study

The Prednisone-Trial randomised a total of 802 adults (Table 1). The total personnel time used for this trial was approximately 2,897 working days and the total costs were calculated to be USD 2.3 million (Table 2; Appendix C). Salary costs and the included overhead expenses accounted for 83.8% (USD 1,928,275) of the total costs. The costs stratified by categories were USD 231,347 (10.1%) for *project development & preparation*, USD 1,938,958 (84.2%) for *patient enrolment, treatment, and follow-up*, and USD 129,518 (5.6%) for the time *after last patient out.*

In the Oxantel-Trial, 480 children were randomised, treated and followed-up within two months in 2012. The total necessary human resources to conduct this RCT were 798 working days and the overall costs for the RCT were USD 100,374 (Table 3; Appendix D). Salary costs accounted for the largest amount of the overall costs (i.e. USD 84,447; 84.1%). Costs for the *project development & preparation* phase were USD 26,437 (26.3%), USD 45,016 (44.8%) for *patient enrolment, treatment, and follow-up*, and USD 28,537 (28.4%) for the *after last patient out* phase.

Table 2: Resource use and associated costs of the Prednisone-Trial.

Cost component	Persons involved	Total time estimation (in days)	Number of items	Costs in 2012 USD	Fraction of total costs (%)
Project development & preparation					
Salary costs:					
Study design, developing study protocol and study documents, incl. analysis plan and source docs	PI, 1 methodological expert, 1 statistician, 13 external experts, 2 residents, 1 data manager	85		81,426	3.5
Preparation for ethical approval at 4 ethics committees, regulatory agency, incl. amendments	PI, 1 resident, 1 study nurse, 10 local investigators (consultants, registrars, and residents), 2 local study nurses	64		38,134	1.7
Set-up meetings with 7 sites, laboratory, and pharmacy	PI, 2 residents, 5 study nurses, 5 laboratory personnel, 2 pharmacy personnel, 10 local investigators	47		37,786	1.6
Other costs:					
Approval fees (ethics committees, Swissmedic)			5	5,433	0.2
Insurance			1	68,568	3.0
Subtotal: Project development & preparation		196		231,347	10.1
Patient enrolment, Treatment, Follow-up					
Salary costs:					
Study drug and matching placebo (production, analytics, packaging, labelling, and randomisation, concealment, etc.)	Pharmacy		800	11,922	3.0
Screening and enrolment at 7 sites	7 residents, 3 registrars, 3 consultants/co-investigators	716		44,7227	19.4
Treatment and follow-up, incl. data entry at 7 sites	8 study nurses, 7 residents, 3 registrars, 3 consultants/co-investigators	1,255		720,848	31.3
Ongoing data management	1 data manager	17		18,144	0.8
Ongoing monitoring	1 monitor	17		14,531	0.6
Ongoing biological samples management, shipping and administration	4 residents, 3 study nurses, 7 laboratory personnel	117		77,687	3.4
AE/SAE documentation, yearly safety report & Data Safety Monitoring Board	PI, 4 residents, 3 study nurses, 6 local investigators (for approval)	67		38,603	1.7
Ongoing meetings of principal investigator	PI	304		292,203	12.7

with staff, collaborations, literature research related to project, etc.					
Ongoing training of staff at sites	PI, 1 resident, 1 study nurse, 5 local investigators, 3 study nurses, 3 residents	40		21,561	0.9
Other costs:					
Study drug compound			400	3,070	0.1
Biological samples (materials, analysis of biological samples)			16,000	218,624	9.5
Other medical material (freezer, blood pressure machine, pulsoxymeters, etc.)			7	23,015	1.0
Other material (flyers, photocopies, folders etc.)				10,549	0.5
Incentives for sites (e.g. breakfasts, dinners, etc.)			38	14,294	0.6
Travel expenses				26,680	1.2
Subtotal: Patient enrolment, Treatment, Follow-up		2,534		1,938,958	84.2
After last patient out					
Salary costs:					
Data cleaning, statistical analysis, and final study report	PI, 2 residents, 1 registrar, 1 statistician, 1 methodological expert, 1 study nurse	86		63,483	2.8
First publication, incl. writing, submission, review process, proofs)	PI, 4 residents, 1 registrar, 1 methodological expert, 1 statistician, 1 secretary	75		59,780	2.6
Conferences	PI, resident	7		4,940	0.2
Other costs:					
Publication fee			1	1,315	0.1
Subtotal: After last patient out		168		129,518	5.6
Additional costs					
Computer rental and software			1	2,061	0.1
Total salary costs				1,928,275	83.8
Total other and additional costs				373,609	16.2
Total		**2,897**		**2,301,884**	**100.0**

Table 3: Resource use and associated costs of the Oxantel-Trial

Cost component	Persons involved	Total time estimation (in days)	Number of items	Costs in 2012 USD	Fraction of total costs (%)
Project development & preparation					
Salary costs:					
Study design, developing study protocol and study documents	1 Principle investigator, 1 PhD student, 2 Statisticians, 3 consulted experts, 2 co-Principal investigators	48.6		20,595	20.5
Production of oxantel pamoate and matching placebo, packaging of study drugs	1 Research Associate (Pharmaceutical Technology), 2 Pharmacists (1 PhD student, 1 PostDoc)	7		3,242	3.2
Other:					
Ingredients for Oxantel tablets				251	0.3
Trial registration and ethical approval from Switzerland and Zanzibar				867	0.8
Insurance for patients				1,482	1.5
Subtotal: Project development & preparation		55.6		26,437	26.3
Patient enrolment, Treatment, and Follow-up					
Salary costs:					
Screening and enrolment	6 Field staff, 6 Laboratory Staff, 2 Data entry staff, 1 cleaner, 2 co-Principal investigators, 1 PhD Student	382		15,873	15.8
Treatment, assessing AE/SAE, and follow-up	2 Local Physicians, 6 Field staff, 6 Laboratory Staff, 1 cleaner, 2 co-Principal investigators, 1 PhD student, 1 Principle investigator, 2 Data entry staff	263		17,850	17.8
Other:					
Material for stool collection (inclusive car) and analysis				3,280	3.3
Transport reimbursement for patients/guardians			600	720	0.7
Study drugs and matching placebos			1,200	497	0.5
Materials and drugs for physicians				370	0.4
Telephone, Fax, Internet, stationery				432	0.4
Costs for investigators from Switzerland (flights, bill of expenses, accommodation)	Principle investigator, PhD student			5,994	6.0
Subtotal: Patient enrolment,		645		45,016	44.8

Treatment, Follow-up					
After last patient out					
Salary costs:					
Treatment of all school-children	6 Field staff, 1 PhD student	7		377	0.4
Data management, analysis, publishing and presenting results	1 Principle investigator, 1 PhD student, 1 Statistician, 2 consulted experts	90		26,509	26.4
Other:					
Drugs for treatment of all school-children			2,100	630	0.6
Conference fee and expenses (British Society for Parasitology)	1 PhD student			1,020	1.0
Publication fees (*New England Journal of Medicine*)				0	0.0
Subtotal: After last patient out		97		28,537	28.4
Additional costs					
Computer rental and software (STATA)				384	0.4
Total Salary costs				84,447	84.1
Total Other and Additional costs				15,927	15.9
Total		**797.6**		**100,374**	**100**

The use of personnel time split by functions is shown in Figure 2. In short, seven residents and five study nurses contributed to 77% of a total of 2,897 working days in the Prednisone-Trial (2,244 days), followed by the principle investigator (408 days,14%), and the five co-investigators (112 days, 4%). All other involved personnel accounted for 1% or less of the total working time.

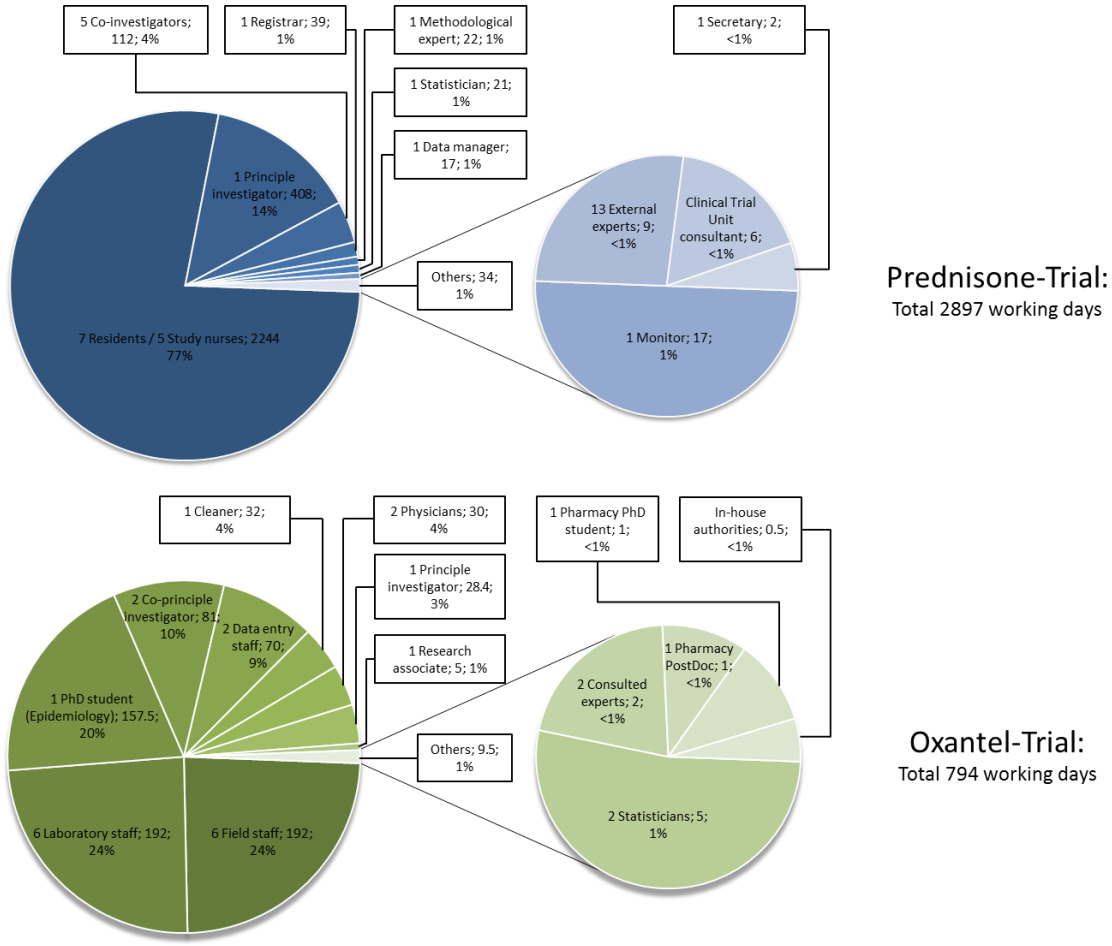


Figure 2: Personnel resources used (Number of involved personnel; number of working days; %) for the Prednisone-Trial and the Oxantel-Trial.

In the Oxantel-Trial, both the six persons working in the field as well as six laboratory personnel contributed 192 working days (24%) each, a PhD student 157.5 (20%) working days, and the two co-principle investigators 81 (10%) days. Two persons conducting data entry worked a total of 70 (9%) days, a cleaner was involved during 32 (4%) days, two physicians worked combined 30 (4%) days and the principle investigator invested 28.4 (3%) days. All other personnel did not account for more than 1% of the total invested time.

The one-way sensitivity analyses demonstrate the influence of alternated cost variables on the total costs of the two RCTs (Table 4). Without the additional 30% of overhead/unproductive time, the total cost would be reduced by 15.9% (total costs USD 1.9 million) for the Prednisone-Trial, and by 11.3% (total costs approximately USD 89,000) for the Oxantel-Trial. An increase in salaries of 50% would increase the total costs of both trials by 25%. Also the impact of the sample size was large for both RCTs: a 50% increase or decrease in sample size resulted in an increase or decrease of the total costs of 32.1% for the Prednisone-Trial and 25.0% for the Oxantel-Trial. A 50% increase of the *project development and preparation* costs or the costs *after last patient out* had a relatively small impact on the Prednisone-Trial (3.4% and 2.8%, respectively) and on the Oxantel-Trial (13.2% and 14.2%, respectively). An increase of 50% for *other costs* (all costs beside salaries) increased the overall costs of both trials by approximately 8%.

Table 4: Results from a series of one-way sensitivity analysis to illustrate the impact of alternated cost parameters on the total costs of the two randomised clinical trials.

Parameter tested	**Costs in 2012 USD (change in %)**	
	Prednisone-Trial	Oxantel-Trial
Baseline	2,301,885	100,374
Salary costs		
No addition of 30% overhead/unproductive time	1,935,739 (-15.9%)	88,999 (-11.3%)
Salaries increase by 30%	2,878,886 (+ 25.1 %)	125,708 (+ 25.2%)
Sample size		
A) Increase in sample size of 50%	3,042,164 (+32.1%)	125,420 (+25.0%)
B) Decrease in sample size of 50%	1,561,605 (-32.1%)	75,327 (-25.0%)
Project development and preparation		
A) Increase of 50% in *project development and preparation* costs	2,380,558 (+3.4%)	113,593 (+13.2%)
B) Decrease of 50% in *project development and preparation* costs	2,223,212(-3.4%)	87,155 (-13.2%)
After last patient out		
A) Costs for after last patient out increase by 50%	2,366,645 (+2.8%)	114,666 (+14.2%)
B) Costs for after last patient out decrease by 50%	2,237,125(- 2.8%)	86,082 (-14.2%)
Other costs		
A) Other costs and additional costs increase by 50%	2,491,160 (+8.2%)	108,337 (+7.9%)
B) Other costs and additional costs decrease by 50%	2,112,610 (-8.2%)	92,411 (-7.9%)

Discussion

Given the lack of empirical evidence on resource use and associated costs of RCTs, we present a comprehensive costing list and detailed data from two RCTs to exemplify how costs and resource use can be systematically assessed. We show that two trials in differing settings had similar cost drivers but resulted in very different total costs (USD 2.3 million vs. USD 100,000). This range of costs is confirmed by our systematic review where we found that published estimates of overall costs for RCTs ranged from USD 209,000 to USD 494 Mio [15, 17] (*Speich et al., submitted together with this manuscript to the JCE*). In line with previous studies [13], personnel costs accounted for the majority of expenditures, mainly during the phase of trial conduct (i.e. *patient enrolment, treatment, and follow-up*). Our sensitivity analysis corroborated the high impact of the personnel costs on the overall costs. It has to be re-emphasized that the personnel costs also include overhead expenditures and hence, infrastructure and some equipment and material (e.g. stationery) is considered indirectly through personnel costs.

Reasons for the differing total costs were manifold. The Oxantel-Trial united many pre-requisites for an inexpensive RCT and stayed within the planned budget. Most of the personnel working time was conducted either by a PhD student or in a low-income country [11]. The diagnostic test for patient screening was easy and inexpensive [18] so that approximately 100 children were screened per day, and due to the high infection rate nearly all of them were eligible to participate in the RCT [9]. Treatment was given on consecutive days, and adverse events were assessed due to the short half-live time [19] until 24 hours after the last treatment. Follow-up was conducted three weeks after treatment.

In contrast, early estimates for the Prednisone trial of USD 1.5 million were exceeded by USD 0.8 Mio mainly due to prolonged recruitment duration by 11 months. Reasons were multifactorial, including increased complexity due to its conduct at seven different hospitals in Switzerland, a high income country with one of the most expensive health care systems including highest salaries [11, 20, 21]. Further, the Prednisone-Trial was conducted over five years, included a larger sample-size, multiple follow-up assessments per patient were necessary, and eligibility and enrolment rates were much lower than expected [10]. It is a common scenario with RCTs that the recruitment rate turns out to be lower than anticipated [6], leading to longer screening, enrolment, and recruitment times, which in turn is associated with higher costs. Underestimation of recruitment rates and lack of feasibility assessment frequently even lead to discontinuation of entire RCTs [6].

The assessed cost data indicate that the *patient enrolment, treatment, and follow-up* phase was in each case the most expensive part (i.e. 84.2% Prednisone-Trial; 46.3% Oxantel-Trial). The percentage difference can be explained by the following reasons: First, the *patient enrolment, treatment, and follow-up* was much shorter in the Oxantel-Trial and, second, this particular phase was conducted in a low-income country. The central role of the *patient enrolment, treatment, and follow-up* phase was additionally highlighted by the one-way sensitivity analysis which indicated that the sample size had a large influence on the overall costs. The sample size, however, affects not only costs during the *patient enrolment, treatment, and follow-up* but also costs during the *project development & preparation* (e.g. insurance fee) and for *after last patient out* (e.g. data cleaning).

A major strength of the current cost evaluation is that, according to our systematic review (*Speich et al., submitted together with this manuscript to the JCE*), this is the first study evaluating empirical costs and resource use of RCTs in the academic setting. While other studies on mainly industry-sponsored trials of all phases exist, they do not offer the same level of detail and transparency as our analysis [1]. We enable readers for the first time to understand how RCT costs actually arose. Moreover, we assessed the working effort for each necessary working step of the RCTs. With these data and the personnel expenses the overall costs of the two RCTs were calculated. In our opinion, this approach reflects the true costs better than taking lump sums which are paid by funders per patient, because these incentives for clinicians are often beyond covering only the expenses [22]. In other cases different funding sources might be combined and therefore the lump sum from a single founder will underestimate the total costs. Furthermore, we used a systematically developed item list to cover all relevant resources and costs of an RCT as comprehensively as possible. However, we realized that the list was not user-friendly enough to efficiently collect costs. In the future, we will therefore continue to adapt and revise the list until it is a tool that may be used by researchers to assess cost data or by clinical investigators to plan and monitor study costs.

On the other hand, our study has the following limitations: First, the majority of the time expenditures in our case study had to be estimated retrospectively within semi-structured interviews. Our approach was similar to the one conduct in three trials [23-25] identified in our systematic literature review (*Speich et al., submitted together with this manuscript to the JCE*). We tried to break down the overall costs into smaller items to assess more accurate data. Furthermore, we added 30% of overhead/unproductive time so that the overall costs of RCTs are not underestimated. However, tools to monitor costs of RCTs are needed to acquire more data with higher accuracy. Second, as our title implies this study is a case

study and cannot be generalised. Both of our included studies were, for example, conducted with relatively cheap drugs. In cases where the intervention itself is more expensive (e.g. cancer drugs) and given multiple times, this factor alone could already lead to RCTs that are several times more expensive than the Prednisone-Trial or the Oxantel-Trial. The highest overall cost of a single RCT that we identified in our systematic review (*Speich et al., submitted together with this manuscript to the JCE*) was reported by Zhu and colleagues with overall costs of USD 454 Mio [26]. Nevertheless, while the economic evaluation of our case study may not be representative for other RCTs, it provides a first indication of cost drivers in the academic setting and hopefully encourages the research community to make their cost and resource use data publicly available.

Tools to budget and manage costs in the academic setting are urgently needed, and we believe that without budgeting and tracking costs efficiently, the clinical research enterprise will stay unsustainable [2] and prone to failure [6]. In addition, further research is needed to investigate cost-efficient solutions for clinical research, such that it becomes and remains affordable for academic investigators. Stakeholders who are able to influence the planning and the design of academic RCTs, such as public funding agencies or research ethics committees, should diligently put emphasis on well-planned *a priori* feasibility assessments and well thought-through budgets.

Conclusion

Evidence on the main cost drivers of clinical research and development is urgently needed. Although our findings may not be transferrable to all academic RCTs, this study proposes a structured and transparent methodology and contributes to a better understanding of the type and range of expenses associated with RCTs. In the future, this model approach may be refined and serve as a toolkit for others. It may allow assessing the costs of RCTs at their institutions retrospectively, monitor costs during the conduct of an RCT on an ongoing basis [27], and developing more comprehensive and reliable budgets before the start of the trial.

The MARTA Group

Benjamin Speich[1], Belinda von Niederhäusern[2], Nadine Schur[3], Lars G. Hemkens[1], Benjamin Kasenda[1,4], Christiane Pauli-Magnus[2], Matthias Schwenkglenks[3,5], Matthias Briel[1,6]

[1] Basel Institute for Clinical Epidemiology and Biostatistics, Department of Clinical Research, University of Basel and University Hospital Basel, Switzerland
[2] Clinical Trial Unit, Department of Clinical Research, University of Basel and University Hospital Basel, Switzerland
[3] Institute of Pharmaceutical Medicine, University of Basel, Basel, Switzerland
[4] Department of Medical Oncology, University of Basel and University Hospital Basel, Switzerland
[5] Epidemiology, Biostatistics and Prevention Institute, University of Zürich, Zürich, Switzerland
[6] Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, ON, Canada

Acknowledgments

We thank Cemile Bathelt, Silke Purschke, Nicole Salvisberg, Said M Ali, Shaali M Ame, and Klaus Ehrlich for their assistance with collecting resource use data.

References

[1] Sertkaya A, Wong HH, Jessup A, Beleche T. Key cost drivers of pharmaceutical clinical trials in the United States. Clin Trials. 2016;13:117-26.
[2] Collier R. Rapidly rising clinical trial costs worry researchers. Cmaj. 2009;180:277-8.
[3] Duley L, Antman K, Arena J, Avezum A, Blumenthal M, Bosch J, et al. Specific barriers to the conduct of randomized trials. Clin Trials. 2008;5:40-8.
[4] Eisenstein EL, Collins R, Cracknell BS, Podesta O, Reid ED, Sandercock P, et al. Sensible approaches for reducing clinical trial costs. Clin Trials. 2008;5:75-84.
[5] Sinha G. European move affects academic trials research. J Natl Cancer Inst. 2006;98:1100-1.
[6] Kasenda B, von Elm E, You J, Blumle A, Tomonaga Y, Saccilotto R, et al. Prevalence, characteristics, and publication of discontinued randomized trials. JAMA. 2014;311:1045-51.
[7] Eisenstein EL. Commentary on Sertkaya et al. and Larson et al. Clin Trials. 2016;13:137-9.
[8] Martin L, Hutchens M, Hawkins C, Radnov A. How much do clinical trials cost? Nat Rev Drug Discov. 2017;16:381-2.
[9] Speich B, Ame SM, Ali SM, Alles R, Huwyler J, Hattendorf J, et al. Oxantel pamoate-albendazole for Trichuris trichiura infection. N Engl J Med. 2014;370:610-20.
[10] Blum CA, Nigro N, Briel M, Schuetz P, Ullmer E, Suter-Widmer I, et al. Adjunct prednisone therapy for patients with community-acquired pneumonia: a multicentre, double-blind, randomised, placebo-controlled trial. Lancet. 2015;385:1511-8.
[11] Bebbington E, Furniss D. Linear regression analysis of Hospital Episode Statistics predicts a large increase in demand for elective hand surgery in England. J Plast Reconstr Aesthet Surg. 2015;68:243-51.
[12] De Tanti A, Zampolini M, Pregno S. Recommendations for clinical practice and research in severe brain injury in intensive rehabilitation: the Italian Consensus Conference. Eur J Phys Rehabil Med. 2015;51:89-103.
[13] Bratt JH, Foreit J, Chen PL, West C, Janowitz B, de Vargas T. A comparison of four approaches for measuring clinician time use. Health Policy Plan. 1999;14:374-81.
[14] Pan J, Zhu JY, Kee HS, Zhang Q, Lu YQ. A review of compression, ventilation, defibrillation, drug treatment, and targeted temperature management in cardiopulmonary resuscitation. Chin Med J. 2015;128:550-4.
[15] Detsky AS. Are clinical trials a cost-effective investment? Jama. 1989;262:1795-800.
[16] Shore BJ, Nasreddine AY, Kocher MS. Overcoming the funding challenge: the cost of randomized controlled trials in the next decade. J Bone Joint Surg Am. 2012;94 Suppl 1:101-6.
[17] Larson GS, Carey C, Grarup J, Hudson F, Sachi K, Vjecha MJ, et al. Lessons learned: Infrastructure development and financial management for large, publicly funded, international trials. Clin. 2016;13:127-36.
[18] Speich B, Knopp S, Mohammed KA, Khamis IS, Rinaldi L, Cringoli G, et al. Comparative cost assessment of the Kato-Katz and FLOTAC techniques for soil-transmitted helminth diagnosis in epidemiological surveys. Parasit Vectors. 2010;3:71.
[19] McCarthy J, Loukas A, Hotez PJ. Goodman and Gilman's: The Pharmacological Basis of Therapeutics, 2nd Edition; Chapter 51 Chemotherapy of Helminth Infections 2011.
[20] De Pietro C, Camenzind P, Sturny I, Crivelli L, Edwards-Garavoglia S, Spranger A, et al. Switzerland: Health System Review. Health Syst Transit. 2015;17:1-288, xix.
[21] de Sa D, Thornley P, Evaniew N, Madden K, Bhandari M, Ghert M. CHAracteristics of research studies that iNfluence practice: a GEneral survey of Canadian orthopaedic Surgeons (CHANGES): a pilot survey. Springerplus. 2015;4:62.
[22] Raftery J, Kerr C, Hawker S, Powell J. Paying clinicians to join clinical trials: a review of guidelines and interview study of trialists. Trials. 2009;10:15.
[23] Evans WK, Dahrouge S, Stapleton J, Quinn C, Pollock D, Waterfield B, et al. An estimate of the cost of conducting phase II trials in lung cancer. Lung Cancer. 2000;28:85-95.

[24] Thornquist MD, Urban N, Tseng A, Edelstein C, Lund B, Omenn GS. Research cost analyses to aid in decision making in the conduct of a large prevention trial, CARET. Carotene and Retinol Efficacy Trial. Control Clin Trials. 1993;14:325-39.
[25] Eri LM, Tveter KJ. Patient recruitment to and cost of a prospective trial of medical treatment for benign prostatic hyperplasia. Eur Urol. 1992;22:9-13.
[26] Zhu CS, Pinsky PF, Kramer BS, Prorok PC, Purdue MP, Berg CD, et al. The prostate, lung, colorectal, and ovarian cancer screening trial and its associated research resource. J Natl Cancer Inst. 2013;105:1684-93.
[27] Eisenstein EL. DRIVE it home: making the case for prospective economic data collection in randomized clinical trials. Kidney Int. 2008;74:1507-9.

Appendix

A – Development of the cost item list

Systematic literature search

A systematic literature search of MEDLINE and EMBASE via the Ovid interphase and EconLit was designed and conducted on Tuesday 3rd November 2015. Our search was not restricted in terms of language, publication type, or publication date.

Ovid search strategy

1 Randomized Controlled Trial* as Topic.mp. or Randomized Controlled Trials as Topic/
2 clinical trial* as topic.mp. or Clinical Trials as Topic/
3 1 or 2
4 (cost* adj2 (component* or item*)).tw.
5 (budget* adj2 calc*).tw.
6 ((study or trial) adj2 budget*).tw.
7 cost* method*.tw.
8 (unit* adj2 cost*).tw.
9 4 or 5 or 6 or 7 or 8
10 3 and 9
11 remove duplicates from 10

EconLit search strategy

1st field: TX All Text:
(randomized control* trial*) or RCT* or (clinical trial*) or (clinical stud*)
AND
2nd field: TX All Text:
(cost* component*) or (cost* item*) or (budget* calc*) or (study budget*) or (trial budget*) or (cost* method*) or (unit* cost*)

Inclusion criteria

At title/abstract level, we included any articles

a. Indicating the stratification of costs related to clinical trials into different costing groups and/or specific cost items.
b. Describing the development and implementation of clinical trial budget templates and budgeting tools.
c. Discussing a single clinical trial-related costing group (e.g. study development, study implementation), as long as this was the overall *topic* of the article according to the abstract/title and this costing group was considered an integral part of the overall study costs.

At full text level, we considered all articles providing specific cost items of clinical trials.

Exclusion criteria

We excluded any articles that focused on i) animal research, ii) hospital / health care costs and iii) cost-benefit analyses.

Study selection

Before starting the systematic literature review, we thoroughly discussed the search strategy, tested the results based on a subset of hits and improved the search criteria in order to obtain more relevant and less irrelevant hits. Afterwards, we performed the search in all databases and screened the articles as outlined above. The number of articles involved in each step is displayed in Figure 1. Whenever studies mentioned websites and/or additional documents provided via the internet, we flagged the articles and checked the information during the systematic internet search process.

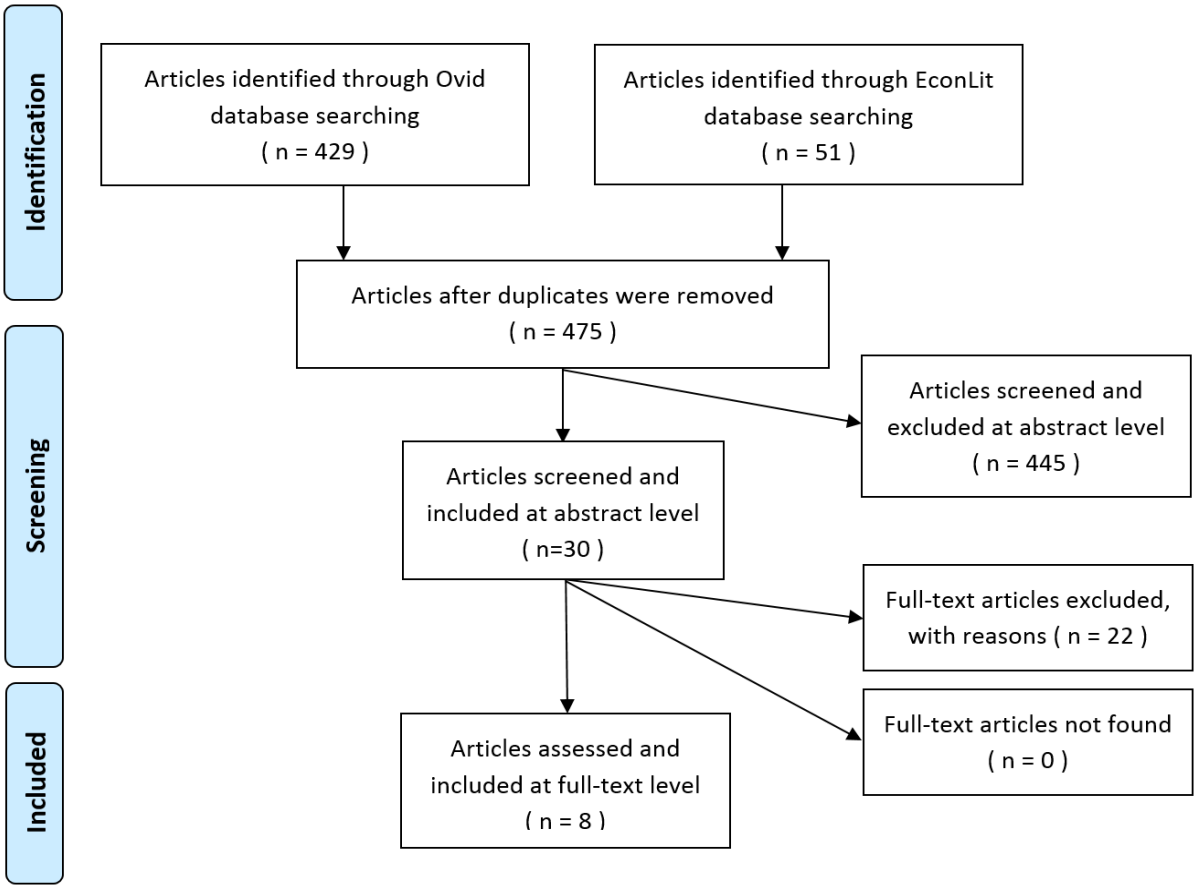


Figure 1: Overview of the systematic literature review process

All relevant cost items related to clinical research trials based on the systematic literature review were gathered and incorporated in an electronic shared database (googledocs.com).

Systematic Internet Search

In addition to the systematic literature review, we systematically searched for direct and indirect cost items associated with all phases of an RCT (i.e. planning/preparation, conduct, analysis, reporting) by systematically screening websites (and any linked information) of various stakeholders involved in clinical research. Stakeholders mostly included governmental bodies, academic research organizations, and funding agencies. Websites were identified through the Google search engine using the keywords “clinical trial”, “randomized controlled trial”, “RCT”, or “clinical research” and “cost”, “costing“, “cost items”, “budget”, “budget template”.

Cost items associated with clinical trials were extracted and compiled in an electronic, shared database (googledocs.com) together with the findings from the literature review. All websites / documents were searched for cost items until no additional items emerged, following the principle of saturation.

In addition to the systematic searches, we received two clinical trial budget templates from institutions involved in clinical research in Switzerland (1 industry / 1 not-for-profit). Cost items were extracted and added until no new items emerged.

B – Cost item list

The cost item list is available as a separate Excel-file.

C – Detailed cost evaluation of the Prednisone trial

Cost component	Persons involved	Total time estimation (in days)	Number of items	Costs in 2012 USD	Fraction of total costs (%)
Project development & preparation					
Salary costs:					
Study design, proposal, documents, analysis plan	PI	52		49,949	2.2
	Methodological expert	4		3,746	0.2
	Statistician	3		2,497	0.1
	External expert	1		2,755	0.1
	External expert	1		2,755	0.1
	External expert	1		2,755	0.1
	10 external experts	3		2,497	0.1
Source Docs	Resident	14		8,142	0.4
Consultation of further experts	Clinical Trial Unit	6		6,329	0.3
After ethics votum: protocol/dossier rewriting	Resident	4		2,443	0.1
Dossier preparation for different Eks	Resident/study nurse	33		18,728	0.8
	PI	3		3,122	0.1
Swissmedic dossier preparation	Resident/study nurse	7		4,071	0.2
Protocol amendments	Resident	17		9,771	0.4
Communication with centres for set-up	PI	16		14,985	0.7
Set-up meeting with pharmacy	Resident/study nurse	3		1,628	0.1
	PI	1		1,249	0.1
Meeting with Laboratories	Resident/study nurse	11		6,514	0.3
	PI	10		9,990	0.4
Set up of sites	Resident/study nurse	6		3,420	0.1
Other costs:					
Approval fees (4 ethics committees, Swissmedic)			5	5,433	0.2
Insurance			1	68,568	3.0
Subtotal: Project development & preparation		196		231,347	10.1
Patient enrolment, Treatment, Follow-up					

Salary costs:					
Study drug and matching placebo (production, analytics, packaging, labelling, and randomisation, concealment, etc.)	Pharmacy	Time missing	800	11,922	0.5
Coordinating site (2328 patients screened, 400 enrolled)					
· Screening	Resident/study nurse	99		57,160	2.5
· Informing participants	Resident/study nurse	164		94,778	4.1
· Enrolment	Resident/study nurse	85		48,855	2.1
Site 2 (458 screened, 40 enrolled)				0	0.0
· Screening	Resident/study nurse	18		10,585	0.5
· Informing participants	Resident/study nurse	32		18,646	0.8
· Enrolment	Resident/study nurse	8		4,885	0.2
Site 3 (96 screened, 33 enrolled)					
· Informing participants	Resident/study nurse	7		3,908	0.2
· Enrolment	Resident/study nurse	7		4,031	0.2
Site 4 (542 screened, 153 enrolled)					
· Screening	Resident/study nurse	46		26,463	1.1
· Informing participants	Resident/study nurse	38		22,066	1.0
· Enrolment	Resident/study nurse	32		18,687	0.8
Site 5 (185 screened, 148 enrolled)					
· Screening	Co-investigator	64		56,817	2.5
· Informing participants	Co-investigator	13		11,551	0.5
· Enrolment	Co-investigator	31		27,722	1.2
Site 6 (11 enrolled)					
· Informing participants	Co-investigator	1		687	0.0
· Enrolment	Co-investigator	2		2,060	0.1
Site 7 (15 enrolled)					
· Informing participants	Resident/study nurse	1		611	0.0
· Enrolment	Resident/study nurse	3		1,832	0.1
Additional time for 1/10 patients for which exclusion criteria was only noticed after informing the patient	Resident/study nurse	6		3,257	0.1
Additional time for 508 patients with immunosuppression	Resident/study nurse	7		4,071	0.2

which were assessed for eligibility					
Additional time for "no ic possible" assessed for eligibility	Resident/study nurse	7		3,925	0.2
Additional time for patients which did not consent	Resident/study nurse	43		24,631	1.1
Treatment / Follow up / Data entry at coordinating site	Study nurse	392		260,460	11.3
Site 2	Study nurse	53		27,684	1.2
Site 3	Study nurse	43		22,840	1.0
Site 4	Study nurse	201		105,892	4.6
Site 5	Study nurse	195		102,432	4.4
Site 6	Study nurse	14		7,613	0.3
Site 7		20		10,382	0.5
Determination of clinical stability	Resident	92		52,926	2.3
Patient discharge	Resident	37		21,496	0.9
Ongoing communication/coordination with participating sites	Resident/study nurse	204		107,350	4.7
Additional insulin therapy necessary in an additional 8 % (n=33) of treatment arm patients:					
Informing patients	Study nurse	3		1,344	0.1
	Resident	1		430	0.0
Insulin (covered by insurance)					
Ongoing data management	Data manager	17		18,144	0.8
Ongoing monitoring	Monitor	17		14,531	0.6
Ongoing sample management	Resident	59		34,153	1.5
Sample sorting (n=16'000)	Study nurse/student	38		25,317	1.1
Shipping: Admin time and fees	Resident/study nurse	20		18,217	0.8
AE/SAE documentation	Resident	13		7,287	0.3
Data Safety & Monitoring Board	External expert	1		936	0.0
	External expert	1		2,755	0.1
Yearly safety report	Resident, study nurse	53		27,626	1.2
Ongoing meetings of principal investigator with staff, collaboratios, literature research related to project, etc.	PI	303		292,203	12.7
Ongoing training of staff at sites	Resident	14		8,142	0.4

	Study nurse	26		13,419	0.6
Other costs:					
Study drug compound			400	3,070	0.1
Biological samples (materials, analysis of biological samples)			16000	218,624	9.5
Other medical material (freezer, blood pressure machine, pulsoxymeters, etc.)			7	23,015	1.0
Other material (flyers, photocopies, folders etc.)				10,549	0.5
Incentives for sites (e.g. breakfasts, dinners, etc.)			38	14,294	0.6
Travel expenses				26,680	1.2
Subtotal: Patient enrolment, Treatment, Follow-up		2,533		1,938,958	84.2
After last patient out					
Salary costs:					
Data cleaning	Resident	28		12,487	0.5
	Registrar	26		12,487	0.5
Statistical analysis	Methodological expert	13		16,285	0.7
	Statistician	13		18,712	0.8
Final study report	PI	1		499	0.0
	Resident	3		1,628	0.1
	Study nurse	3		1,384	0.1
Writing of main publication (Lancet)	PI	9		8,741	0.4
	Resident	14		8,142	0.4
	Registrar	7		4,678	0.2
	Methodological expert	3		2,497	0.1
	Statistician	3		2,497	0.1
Submission procedures	Resident	1		814	0.0
	Secretary	2		814	0.0
Review process	PI	8		8,991	0.4
	Resident	10		5,863	0.3
	Methodological expert	3		2,497	0.1
	Statistician	3		2,497	0.1
	Registrar	7		6,243	0.3
	External expert	1		1,377	0.1

Proofs	PI	3		2,497	0.1
	Resident	3		1,628	0.1
Other costs:					
Publication fee				1,315	0.1
Conference fee and expenses	Resident	7		4,940	0.2
Subtotal: After last patient out		168		129,518	5.6
Additional costs					
Computer rental and software				2,061	0.1
Total costs		**2,897**		**2,301,884**	**100.0**

D - Detailed cost evaluation of the Oxantel trial

Cost component	Persons involved	Total time estimation (in days)	Number of items	Costs in 2012 USD	Fraction of total costs in %
Project development & preparation					
Salary costs:					
Preparation of research protocol (e.g. literature review, development of idea, writing-up, meeting time)	1 Principle Investigator 1 PhD Student 2 Statistician 2 Consulted experts 1 In house authority	7.8 15.6 1 1 0.5		5,343 3,637 844 1,055 527	5.3 3.6 0.8 1.1 0.5
Acquiring funding	1 Principle Investigator	6.5		4,453	4.4
Developing/updating investigational brochure, standard operating Procedures, case report form and patient related forms	1 Principle investigator 1 PhD Student 1 Co-Principle investigator 1 In house authority	0.25 7.2 1 0.25		171 1,679 60 264	0.2 1.7 0.1 0.3
Applications to ethics committee, authorities, insurance (e.g. preparation, submission; but not fees)	1 Principle investigator 1 PhD Student	0.5 3.4		343 793	0.3 0.8
Budgeting (study budget calculation and controlling)	1 PhD Student	1.3		303	0.3
Reporting to funders, managing clinical trial portals	1 Principle investigator 1 PhD Student	1 1		891 233	0.9 0.2
Production of Oxantel Pamoate and matching placebos	1 Research Associate (Pharmaceutical Technology)	5		2,637	2.6
Packaging and labelling of study drugs	1 Pharmacy PostDoc 1 Pharmacy PhD student	1 1		356 249	0.4 0.3
Other costs:					
Ingredients Oxantel Pamoate			1	251	0.3
Ethical approval in Basel, Switzerland			1	317	0.3
Ethical approval on Pemba Island, Tanzania			1	240	0.2
Trial registration			1	310	0.3

Insurance for participants			1	1,482	1.5
Subtotal: Project development & preparation		55.6		26,437	26.3
Patient enrolment, Treatment, Follow-up					
Salary costs:					
Screening and enrolment	6 Field staff 6 Laboratory staff 1 Cleaner 2 Co-Principle investigator 2 Data entry staff 1 PhD student	120 120 21 48 42 25		2,880 3,024 252 2,880 1,008 5,829	2.9 3.0 0.3 2.9 1.0 5.8
Treatment, assessing AE/SAE, and follow-up	2 Local Physicians 6 Field staff 6 Laboratory Staff 1 Cleaner 2 Co-Principle investigator 2 Data entry staff 1 PhD student 1 Principle investigator	30 66 66 11 32 28 23 7		1,800 1,584 1,584 132 1,920 672 5,363 4,795	1.8 1.6 1.6 0.1 1.9 0.7 5.3 4.8
Other costs:					
Material for stool collection (stool containers)			2,400	240	0.2
Material for stool analysis (Kato-Katz Kit, wire mesh)				1,072	1.1
Hired car (in days) including fuel and driver			40	1,920	1.9
Transport reimbursement for patients/guardians			600	720	0.7
Albendazole			300	90	0.1
Mebendazole			150	38	<0.1
Albendazole matching placebo			300	158	0.2
Mebendazole matching placebo			450	211	0.2
Materials for clinical examination and adverse event mitigation				370	0.4
Telephone, Fax, Internet, stationery				432	0.4
Costs for 2 investigators from Switzerland (flights, bill of	1 Principle investigator, 1 PhD student			5,994	6.0

expenses, accommodation)					
Subtotal: Patient enrolment, Treatment, Follow-up		645		44,968	44.8
After last patient out					
Salary costs:					
Treatment of all school-children	6 Field staff 1 PhD student	6 1		145 232	0.1 0.2
Data management, data cleaning	1 PhD student	4		933	0.9
Statistical analyses	1 PhD student 1 Statistician	6.5 1		1,516 844	1.5 0.8
Output Tables and Figures plus interpretation of results	1 PhD student 1 Principle investigator	13 0.5		3,031 343	3.0 0.3
Manuscript preparation and revision	1 PhD student 1 Principle investigator 1 Statistician 2 Consulted experts	53 4.5 3 1		12,358 3,083 1,532 1,055	12.3 3.1 1.5 1.1
Conference (including abstract submission)	1 PhD student	3.5		816	0.8
Other costs:					
Albendazole for treatment of all school-children			2100	630	0.6
Hired car (in days) for treatment of all school children			1	48	0.1
Conference fee and expenses (British Society for Parasitology)	1 PhD student		1	1,020	1.0
Publication fees (New England Journal of Medicine)	1		1	0	0.0
Subtotal: After last patient out		97		28,585	28.5
Additional costs					
Computer rental and software (STATA)	2			384	0.4
Total costs		**797.6**		**100,374**	**100.0**

Manuscript V: Generating evidence on a risk-based monitoring approach in the academic setting – Lessons learned

Belinda von Niederhäusern[1*]; Annette Orleth[2], PhD; Sabine Schädelin[1]; Nawal Rawi[3]; Martin Velkopolszky[1]; Claudia Becherer[1], Pascal Benkert[1], PhD; Priya Satalkar[4,5], PhD; Matthias Briel[5,6], MD; Christiane Pauli-Magnus[1], MD

[1]Clinical Trial Unit, Department of Clinical Research, University and University Hospital Basel, Basel, Switzerland (Emails: belinda.vonniederhausern@usb.ch; sabine.schaedelin@usb.ch, martin.velkopolszky@usb.ch; pascal.benkert@usb.ch; christiane.pauli-magnus@usb.ch)
[2]Department of Medicine, Biomedicine and Clinical Research, Neurology, University Hospital Basel, Basel, Switzerland (Email: AnnetteNicole.Orleth@usb.ch)
[3]Novartis Pharma, Basel, Switzerland (Emails: nawal.rawi@novartis.com)
[4]Institute for Biomedical Ethics, University of Basel, Basel Switzerland (priya.satalkar@unibas.ch)
[5]Basel Institute for Clinical Epidemiology and Biostatistics, Department of Clinical Research, University and University Hospital Basel, Basel, Switzerland (Emails: Matthias.briel@usb.ch)
[6]Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Canada (Emails: Matthias.briel@usb.ch)

Status: Published
BMC Med Res Methodol. 2017 Feb 14;17(1):26.

Abstract

Background: In spite of efforts to employ risk-based strategies to increase monitoring efficiency in the academic setting, empirical evidence on their effectiveness remains sparse. This mixed-methods study aimed to evaluate the risk-based on-site monitoring approach currently followed at our academic institution.

Methods: We selected all studies monitored by the Clinical Trial Unit (CTU) according to Risk ADApted MONitoring (ADAMON) at the University Hospital Basel, Switzerland, between 01.01.2012 and 31.12.2014. We extracted study characteristics and monitoring information from the CTU Enterprise Resource Management system and from monitoring reports of all selected studies. We summarized the data descriptively. Additionally, we conducted semi-structured interviews with the three current CTU monitors.

Results: During the observation period, a total of 214 monitoring visits were conducted in 43 studies resulting in 2961 documented monitoring findings. Our risk-based approach predominantly identified administrative (46.2%) and patient right findings (49.1%). We identified observational study design, high ADAMON risk category, industry sponsorship, the presence of an electronic database, experienced site staff, and inclusion of vulnerable study population to be factors associated with lower numbers of findings. The monitors understand the positive aspects of a risk-based approach but fear missing systematic errors due to the low frequency of visits.

Conclusions: We show that the factors mostly increasing the risk for on-site monitoring findings are underrepresented in the current risk analysis scheme. Our risk-based on-site approach should further be complemented by centralized data checks, allowing monitors to transform their role towards partners for overall trial quality, and success.

Background

Adherence to the International Conference on Harmonization of Good Clinical Practice (ICH GCP) guidelines should ensure the safety, rights, and integrity of trial participants as well as the confidentiality of personal information and data quality [1]. Trial monitoring through trained clinical monitors is requested by ICH GCP, but the guideline provides limited insight on the procedures of quality assessment during such monitoring visits [2, 3]. Traditional approaches relied on intensive on-site visits and 100% source data verification (SDV) irrespective of the risk levels in the study, which have been associated with high cost and limited contribution to clinical trial data quality [4-6].

Recent developments at international bodies and regulatory agencies such as the European Medicines Agency (EMA) have supported the need for risk-proportionate approaches to clinical trial monitoring [7-9]. In November 2016, the ICH published the final version of the integrated addendum to ICH-GCP, advising Sponsors to develop a systematic, prioritized, risk-based approach to monitoring clinical trials [3]. Similarly, the forthcoming European Union (EU) Clinical Trial Regulation will permit reduced monitoring for low-risk intervention trials [10]. Among the first, the Risk ADApted MONitoring (ADAMON) Project proposed an instrument for the facilitation of risk analysis allowing on-site monitoring strategy tailored to the risk profile of every trial [11]. Risk analysis thereby refers to the risk of jeopardizing patient safety and rights or the validity of results and considers patient, site, and study design robustness-related indicators. Furthermore, risk analysis takes into account the risks of the study intervention compared to the risks a patient would run if treated in routine practice. This approach was first proposed in 2009 and later adapted by other stakeholders such as the Organization for Economic Co-operation and Development (OECD), the U.S. Food and Drug Agency (FDA), and EMA [9, 12, 13]. It encouraged study sponsors to assess, on a case-per-case basis, the risk associated with an individual trial protocol, implement risk assessments that focus on critical data and procedures, and utilize alternative monitoring approaches taking advantage of the increasing use of electronic systems.

Sponsors should develop a monitoring plan that describes, based on the risk assessment, the monitoring strategy, the monitoring responsibilities of all the parties involved, the various monitoring methods to be used, and the rationale for their use [3]. However, in the absence of credible data to describe impact of a change of monitoring approach on data quality and study cost, the majority of industry-sponsored trials continue to be monitored using a traditional monitoring approach with up to 100% SDV. It has been estimated that SDV can consume up to 25% of the sponsor’s entire clinical trial budget, even though the association between data quality/subject safety and the extent of monitoring and SDV has not been

clearly demonstrated [14]. Financial estimates of a single monitoring site visit range from US$800 in 1991 to US$1500 in 2009 [15, 16], with conservative cost estimates for one single query of US$150 [17]. The approach taken may therefore be evaluated as overcautious at best, and at worst, a complete waste of resources based on current reviews [18, 19].

In the academic setting, restricted resources often oblige investigators to apply a risk-based approach to trial monitoring which is expected to be less labor intense [20]. At the academic Clinical Trial Unit (CTU) at the University Hospital in Basel, Switzerland, we have applied risk-based on-site monitoring based on the ADAMON project for all patient-oriented research projects since 2012. In order to understand the implications of this approach for patient safety and data quality at our institution, we undertook this mixed-method investigation. The aim of our study was to i) retrospectively investigate the characteristics of findings documented during on-site visits, ii) identify key factors that might influence the number and types of monitoring findings, iii) assess the costs associated with our approach, and iv) understand the experience of our monitors and the challenges they face.

Methods

Setting

This mixed-method study was performed at the CTU of the University Hospital in Basel. The CTU offers monitoring services to investigator- and industry-initiated studies conducted at our institution and affiliated sites if desired by sponsors. CTU monitors are qualified by training and experience and work according to clearly defined standard operating procedures (SOPs) which are reviewed and updated by an autonomous quality-assurance officer on a regular basis. The risk evaluation adopted by the CTU (*Table 1*) includes a structured trial risk classification by the project manager according to the ADAMON project and the Swiss Human Research Act as described by the Swiss Clinical Trial Organization [21]. This approach allows the categories low, medium, or high risk; and the assessment of additional three risk modulators (*Table 1*). These risk modulators may lower or raise the risk within a certain risk category and therefore influence the duration of site visits, but not their frequency. After risk classification, the project manager specifies the extent and nature of on-site monitoring visits in the monitoring plan *(Table 2).* CTU monitors then conduct on-site monitoring visits according to the pre-specified monitoring plan and document monitoring findings in monitoring reports which are shared and discussed with both the sponsor and the project manager.

Table 1. Risk classification procedure at Clinical Trial Unit Basel. 1-3 conducted according to the Swiss Clinical Trial Organization Guidelines for Good Operational Practice V2.0. Scheme adapted from Hurley et al. [22].

Risk classification procedure				**Recommended systematic review of trial's risk profile**
1. Initial Risk classification	**2. Categories of risk modulators**	**3. Risk classification**	**4. Modulators of monitoring extent**	
Potential risk of therapeutic intervention in comparison to standard of medical care (as described in HRA & ADAMON): • Comparable (see also ClinO art. 19,20,61, category A) • Higher (see also ClinO art. 19,61, category B) • Markedly Higher (see also ClinO art. 19,20, category C)	Modulators from the following ADAMON categories may influence the initial risk category by a max. of +1 or -1 risk category (e.g. Intermediate to High Risk): 1. Potential trial participant-related critical indicators[1] 2. Robustness related indicators – "hard primary endpoints" and/or simple clinical trial procedure 3. Site-related indicators[2]	An overall risk category is assigned based on the results of 1) and 2) as follows: • Low risk: Risk of intervention comparable & trial has at least one indicator of robustness and no participant related indicator • Intermediate risk: Risk of intervention comparable, or higher & trial has no indicator of robustness, or at least one participant-related indicator present • High risk: Risk of intervention higher or markedly higher, and trial has no indicator of robustness, or at least one participant related critical indicator present	The following modulators may influence monitoring extent (i.e. number of hours per visit) within risk category: 1. site experience with clinical trials 2. presence of an electronic database at site 3. whether site is coordinating lead site of the study	An additional risk assessment is required if the trial undergoes substantial amendments

[1] Including indicators on vulnerability of study population, setting of recruitment, critical eligibility criteria, additional risks of therapy, trial procedures that are unusually complex, etc. [2] Including essential technical, personnel, storage, transport, or documentation requirements at site. Site-related indicators do not affect the risk category of a study, but may modulate the extent and duration of individual monitoring visits. Human Research Act (HRA), ADApted MONitoring (ADAMON), Clinical Trials Ordinance (ClinO)

Table 2. Recommended on-site monitoring activities based on study risk classification. Informed Consent (IC)

Risk of Study	Initiation visit	Interim visit	Content of interim visits	Close out visit
Low	optional	after first patients, then adaptable (e.g. 1/year)	Endpoints (extent to be defined), IC (usually 100%)	optional
Intermediate	mandatory	after first patients, then adaptable (e.g. 1/year)	Endpoints (extent to be defined), IC (usually 100%), safety (usually 100%)	mandatory
High	mandatory	after first patients, then in regular intervals	Endpoints (extent to be defined), IC (usually 100%), safety (usually 100%)	mandatory

Quantitative retrospective analysis

We included all investigator-initiated trials (IITs) and industry-sponsored studies monitored by the CTU between January 1st 2012 and December 31st 2014 with the exception of studies for which monitoring had never been fully initiated (i.e. <10% of planned working hours completed because of an early study discontinuation or delayed study start). Since the introduction of risk-based monitoring at our institution in 2012, a total of six monitors had been involved in monitoring activities. For all included studies, we extracted a set of variables covering detailed characteristics at the level of the study itself, the level of the study site and the individual monitoring visit.

Study-specific variables included

- study design,
- study type,
- study sponsor,
- type of research,
- study phase (I-IV), and
- type of study population (e.g. inclusion of vulnerable populations).

Variables covering study site information included

- site location,
- ADAMON risk category,
- presence of electronic database,
- principal investigator, and whether he/she changed during conduct,
- staff experience, and
- number of planned patients at the site.

At the level of the individual monitoring visit we extracted information on

- type of visit (i.e. initiation, interim, close-out),

- the number of
    - administrative,
    - patient rights,
    - patient safety,
    - laboratory/biological specimen,
    - data point confirmation, and
    - endpoint related findings.

Extraction and categorization of findings was performed independently and in duplicate (AO, MV, CB) from monitoring plans and reports using a validated web-based database (secuTrial®). Classification of findings corresponded to the main categories used in our monitoring reports and categories were treated as mutually exclusive. Table 6 provides examples of findings for each category. Discrepancies between extractors in classifying the variables were resolved through discussion by the extractors. After an initial calibration phase, agreement between the extractors was considered "good" if no more than 4 out of 49 extracted variables differed. Findings that were corrected immediately on-site were often not documented and therefore not included in our study. We summarized the number of findings descriptively, stratified by key variables and graphically displayed as i) total findings per study (or site, depending on the variable), ii) percentage of administrative or patient right findings out of total number of findings. Figures were interpreted visually. Furthermore, we collected information on human and financial resources employed for monitoring activities from the CTU Enterprise Resource Management (ERP) system for each project. We calculated total resource use by summing the total hours worked by our monitors during the analyzed time period (2012-2014), as retrieved from the ERP, multiplied by the hourly salary rate. We then divided the total human resource cost by total number of findings which we had documented and extracted from monitoring reports.

Semi-structured interviews

We interviewed three monitors who were involved in these monitoring visits and who continue to work at our institution at present. The main themes covered during these interviews were i) monitors' perspective on risk-based monitoring per se, ii) the practical settings in which these visits and findings of events were documented, iii) the challenges they faced during these visits, and iv) their perspectives on the future development of risk-based monitoring. As interviews did not include health-related data and were therefore not within the scope of the applicable Human Research Act (HRA, Art.1), we did not require formal ethical approval. Each interview was conducted in German by NR, tape recorded with the monitor's permission, transcribed in full, and anonymized at the level of transcription. We

examined all the transcripts in duplicate (BvN and NR) and BvN coded each interview. We then grouped codes into clusters around similar and interrelated themes until we reached consensus. In the results section below, we will describe and discuss three key themes that emerged from our qualitative interviews *(a)* factors influencing risk-based monitoring findings; (b) the monitoring process and the challenges faced**;** and (c) the current role of monitors and future perspectives.

Results

Study sample characteristics

We included forty-three studies (39 investigator-initiated, 3 industry-sponsored) monitored between January 1st 2012 and December 31st 2014 for analysis. Characteristics of these studies are shown in *Table 3*, study stratification by risk categories and associated risk factors in *Table 4.*

Table 3. Study sample characteristics (number, %). Study sample including 43 studies monitored by the CTU Basel between 2012 and 2014.

			Total
		n	%
Total studies		43	100
Study design	Interventional	34	79.1
	Observational	9	20.9
Study type	Multicenter	10	23.3
	Singlecenter	33	76.7
Study sponsor	Investigator (academic)	40	93.0
	Industry	3	7.0
Type of research	Drug	29	67.4
	Medical Device	5	11.6
	Biological Samples[1]	4	9.4
	Other[2]	5	11.6
Study phase (drug studies, n=29)	I	9	31.1
	II	7	24.1
	III	8	27.6
	IV	3	10.3
	Other[3]	2	6.9

[1] Biological samples incl. physiological or genetic analysis of human biological samples (e.g. urine, blood, tissue, etc.)
[2] Other incl. observational research, health economics assessments, or tissue-based intervention/stem-cell transplantation
[3] Other incl. cost-effectiveness trials not specific to a phase

Table 4. Study sample by risk categories and associated risk factors. Study sample including 43 studies monitored by the CTU Basel between 2012 and 2014, stratified by ADAMON risk categories, and factors associated with risk evaluation.

		Total	
		n	%
Total studies		43	100
ADAMON risk category	Low	11	25.6
	medium	23	53.5
	High	9	20.9
	Total	43	100
Electronic database present at first patient in	Yes	19	44.2
	No	24	55.8
	Total	43	100
Principal Investigator change during study	Yes	3	7.0
	No	40	93.0
	Total	43	100
Vulnerable study population[1]	Yes	7	16.3
	No	36	83.7
	Total	43	100
Total sites		94	100
Staff experienced[2]**, by site**	Yes	88	93.6
	No	6	6.4
	Total	94	100
Staff change, by site	Yes	11	11.7
	No	48	51.1
	Unknown	35	37.2
	Total	94	100

[1] Defined as "children, adolescents, adults lacking capacity in the consent procedure, pregnant women and in-vitro fertilized embryos and fetuses, prisoners, and subjects in emergency situations" (according to HRA, Chapter 3).
[2] Defined as a) GCP trained, b) solely dedicated to research activities (e.g. a study nurse, resident, etc.), and c) has been involved in the conduct of one or more clinical research studies before.

Characteristics of monitoring findings

In total, we documented 2961 findings during 214 monitoring visits in 43 studies between 2012 and 2014 (*Table 5, 6*). In ten out of 43 studies, we monitored more than one site. Overall, administrative findings (46.2%; e.g. missing CVs or incomplete Investigator Site Files etc.) were equally predominant as patient rights findings (49.1%; e.g. wrongly signed and dated informed consent forms), whilst patient safety issues were found only

exceptionally (1.1%). Although the studies varied in their total amount of findings, we documented at least one administrative and one patient right finding in almost every study, and at least one safety finding in a fifth of all studies (*Table 5, 6*). The remaining findings included issues related to laboratory procedures or biological specimen (2.3 %), and issues related to the endpoint which could not be clarified with staff at site and required written confirmation (e.g. clarification of a questionable laboratory value which seemed out of range, 1.2%) (*Table 5*).

Table 5. Characteristics of monitoring findings.
Sum of findings by CTU monitors in total (number, %). Note: Findings which were resolved on-site between monitor and study staff and not documented in monitoring reports are not listed.

		Total	
		n	%
Total Studies		43	100
Total Monitoring Visits		214	100
Findings			
	Administrative	1367	46.2
	Patient rights	1453	49.1
	Patient safety	32	1.1
	Laboratory/biol. specimen	70	2.3
	Endpoint related data point: confirmation requested	36	1.2
	Endpoint related data point: Data point changed	3	0.1
	Total	2961	100
Average n findings/visit		13.8	
Studies with			
	at least 1 administrative finding	43	100
	at least 1 patient right finding	41	95.3
	at least 1 patient safety finding	9	20.9

Table 6. Examples of monitoring findings. Curriculum Vitae (CV), Case Report Form (CRF), Ethics Committee (EC)

Finding category	Examples of findings
Administrative	• Changes at the investigational site (staff training, staff CVs, address, technical equipment, etc.) not documented • Functions and responsibilities log not up to date • Subject related logs not up to date • CRFs not available at site and/or not documented by authorized staff
Patient rights	• Informed Consent Forms not signed and/or not dated correctly • No valid and approved version of Informed Consent Form used • Amendments/addenda to Informed Consent Form not communicated to patients and no re-consent obtained • Patient did not fulfill all inclusion criteria
Patient safety	• No description of the process for detecting and reporting serious and unexpected adverse events and/or unanticipated problems involving risk to participants in place at site • Adverse events not correctly documented and/or reported as required (e.g. to Sponsor, EC, Competent Authority) • New safety information not approved by authorities • Staff not trained according to new safety information
Laboratory /Biological Specimen	• Biological specimen not stored correctly according to protocol • Process conducted not in accordance with Good Manufacturing Practice (GMP)
Data point confirmation requested	• Indicates whether finding challenges the credibility of data point, e.g. by stating “Please confirm that blood pressure measure is 100/65mmHg”
Data point changed	• Indicates whether data point was adjusted as direct consequence of finding, e.g. “Blood pressure of 100/65mmHg was corrected to 120/80mmHg”

Influencing factors on number and type of findings

Generally, the sample size of a study was positively associated with the total number of findings (*Figure 1*). Due to the low number of other than administrative and patient rights findings, the figures display both the percentage of patient rights (X %) and administrative findings (100%-X %). Visual inspection of figures showed that factors such as observational study design (*Figure 2a/b*), high ADAMON risk category (*Figure 3b*), industry sponsorship (*Figure 3c*), the presence of an electronic database (*Figure 3d*), experienced site staff (*Figure 3e*), and inclusion of vulnerable study population (*Figure 3f*) were associated with lower numbers of monitoring findings. As a trend, studies sponsored by industry or with a high risk category tended to result in less patient rights findings compared to other studies, for which the proportion pattern (patient rights vs. administrative) varied widely (*Figures 3b and c*).

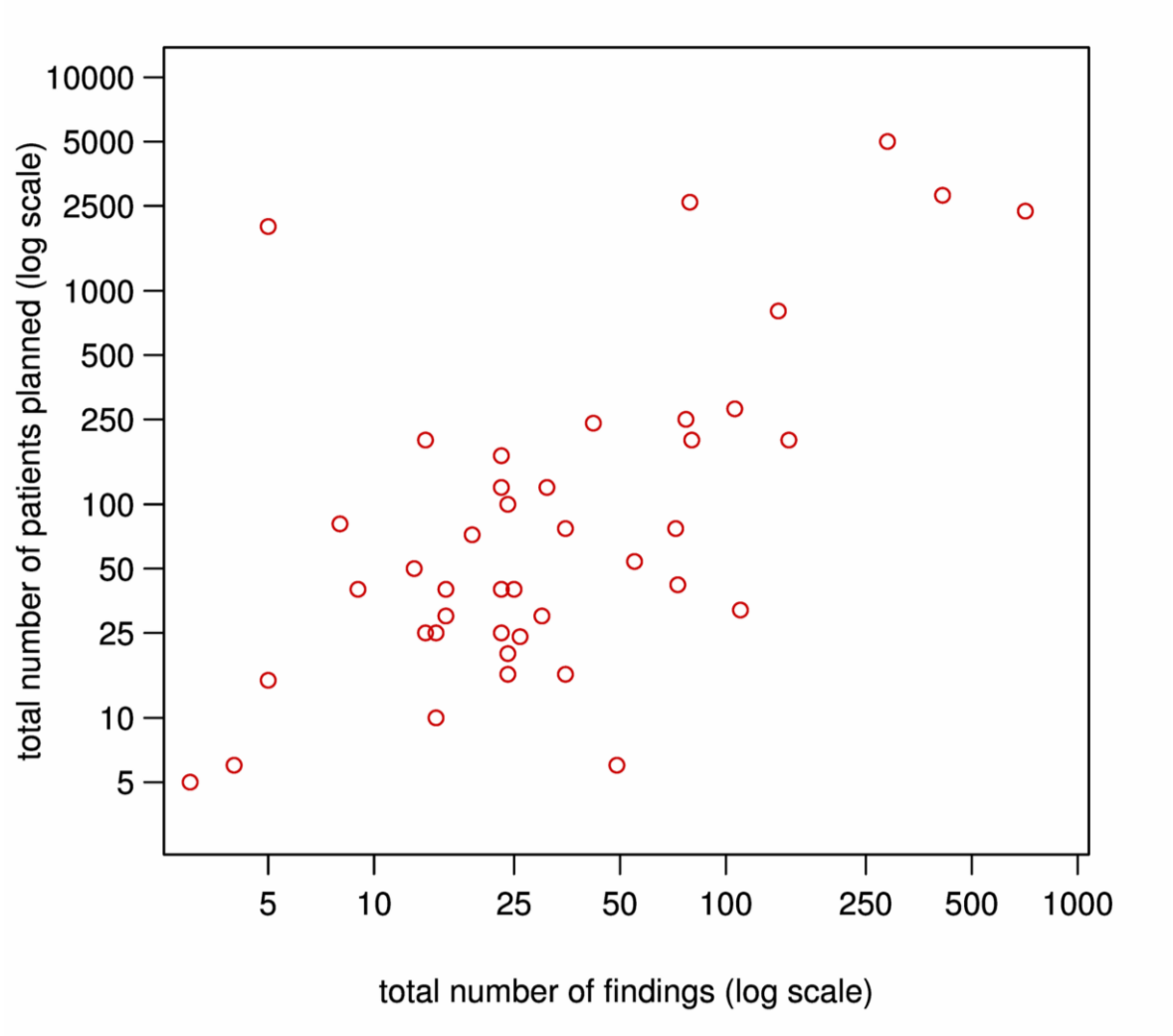
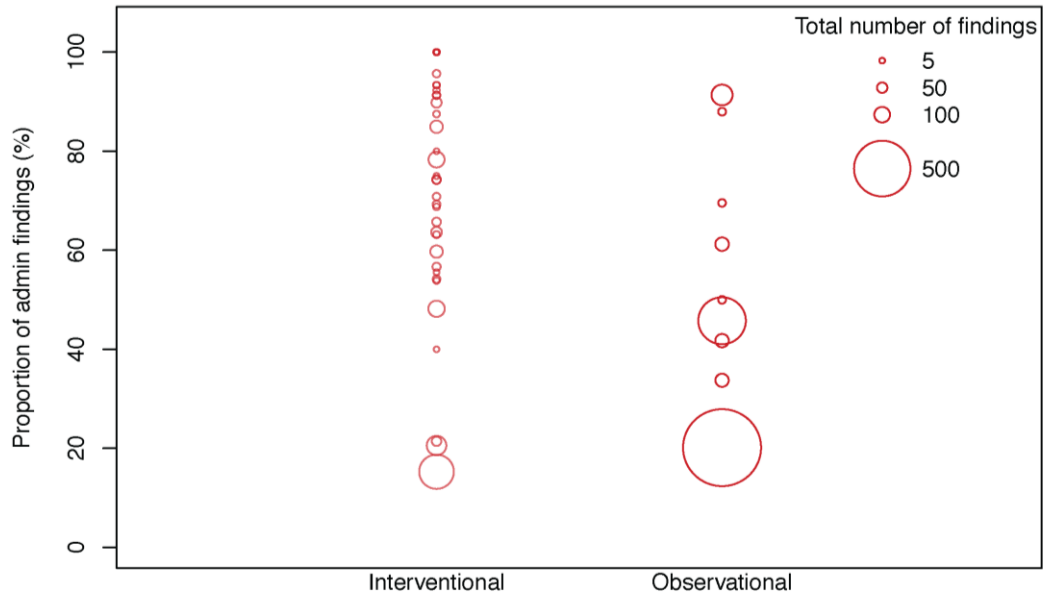


Figure 1. Studies according to the planned sample size and the final total number of monitoring findings (log scale).



b

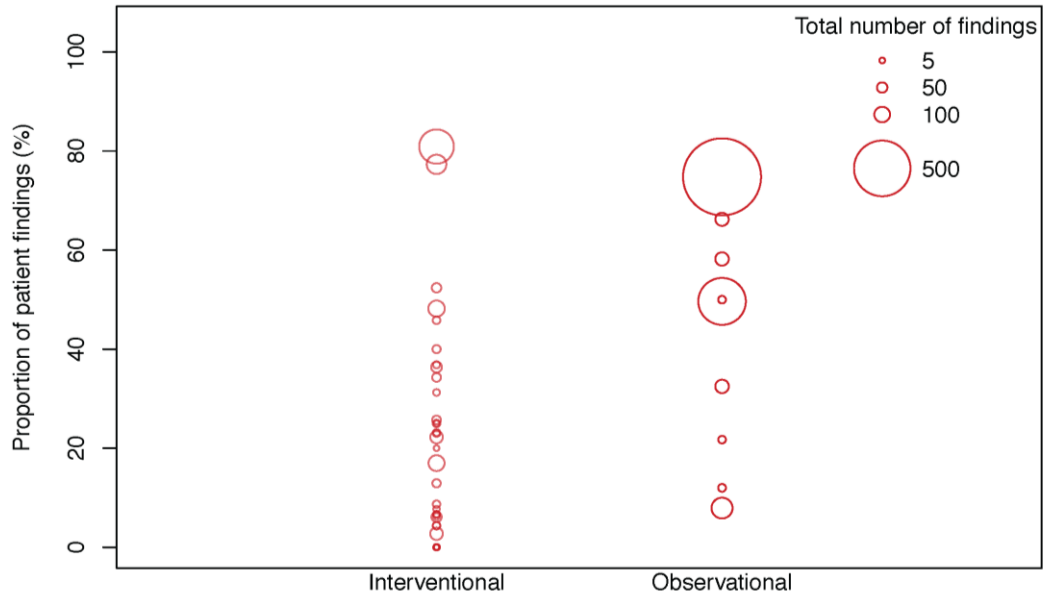


Figure 2. Total number of findings and proportion of administrative (a) and patient rights (b) findings in interventional and observational studies, by study. Diameter of circles proportionate to total number of findings per study.

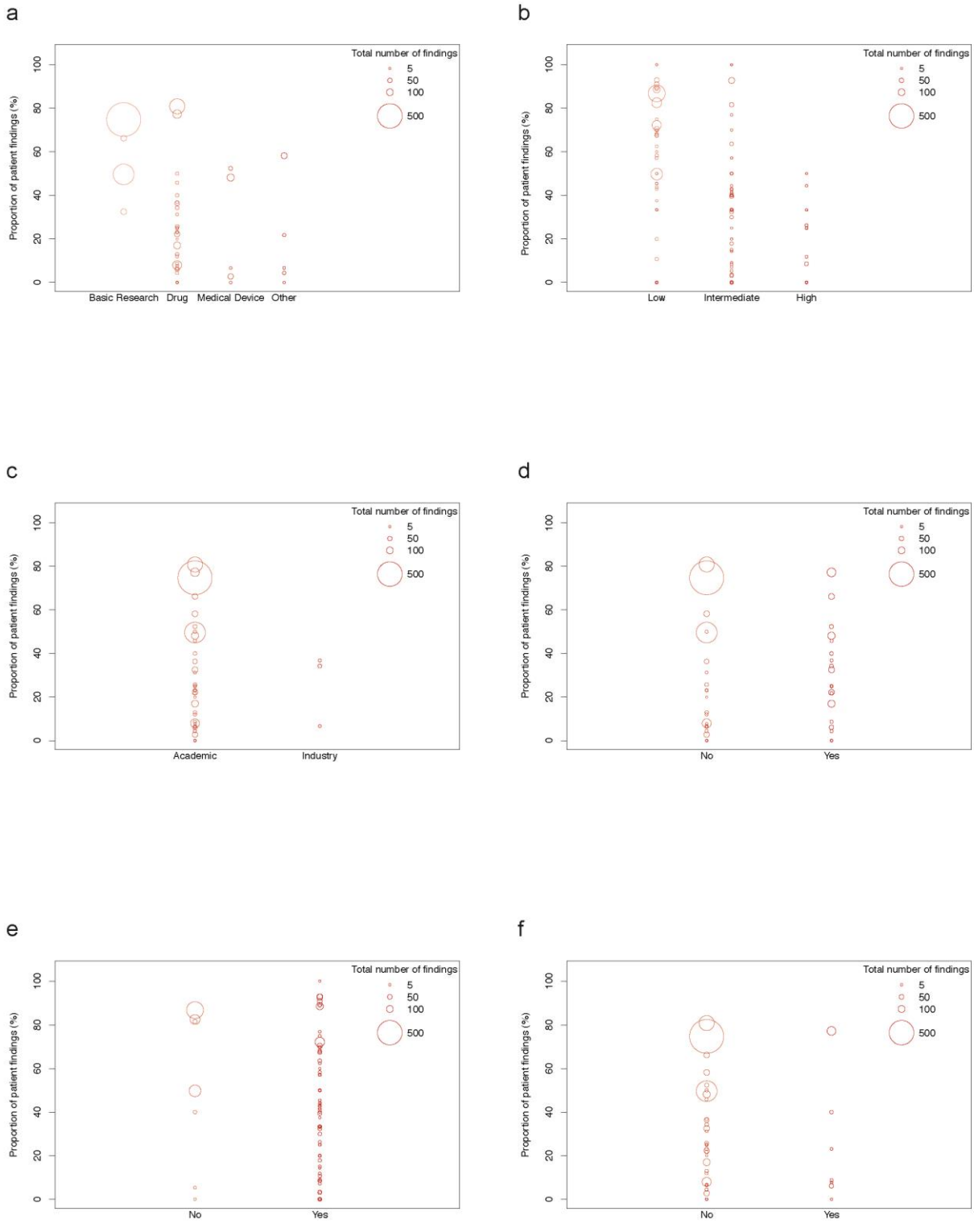


Figure 3. Total number of findings and proportion of patient rights findings in studies stratified by (a) study type, (b) ADAMON risk category, (c) study sponsor, (d) studies with vs. without electronic database, (e) studies conducted at sites with vs. without clinical research experience, (f) studies with vs. without vulnerable study population. Diameter of circles proportionate to total number of findings per study (a, c, d, f) or per site (b, e).

Although observational studies generally resulted in fewer findings, two of the nine analyzed studies were outliers (> 400 findings/study) (*Figure 2a/b*). One was a multicenter study including 7 sites but no electronic data capture system, resulting in a total of 413 findings (45.8% administrative, 49.6% patient rights) in seven initiation, seven interim 1, and five interim 2 visits. The second study was a large single center study (> 2000 planned patients) with inexperienced study personnel and principal investigator. No initiation visit was performed given the low risk character of the study and an electronic data capture system was not available. In this study, four interim visits resulted in a total of 710 findings of which 20.1% were administrative and 75% were related to patient rights issues (*Figure 2b, largest red circle*). In addition, both studies experienced a change in monitor, after which the overall number of findings increased.

Out of 43 monitored studies, 39 were monitored more than once (at least one site), and 12 were monitored at least three times (at least one site) (*Appendix Figure 1a*). Thereof, 11 studies had an initiation visit and four studies experienced a change in monitor throughout the study. Generally, findings tended to decrease after the second interim visit. One study increased in findings after the third visit which was due to a new version of the informed consent which was not adequately used in all patients. The proportion of administrative findings was high at initiation but showed a decrease during the conduct of the study, whereas patient rights findings increased. (*Appendix Figure 1b/c*).

In our sample, only three sites conducted three or more different studies within the given two-year time period (*Appendix Figure 2*; site 1: studies 3, 4, 8, and 9; site 2: studies 22, 23, 33, and 36; site 3: studies 26, 21, and 32). Factors such as study design (interventional vs. observational), study type (e.g. phase 1-3), sample size, the risk associated with the studies performed (and therefore the associated monitoring risk category), and staff experience does usually not differ much within a given site, and visual inspection did not reveal a major difference in total number of findings within trials at each of the three sites (*Appendix Figure 2*). However, the mentioned characteristics differed significantly across the three sites (one high risk pharmacological phase 1 unit, one low risk observational cardiology unit, and one medium risk cognitive neuroscience unit) and did therefore not allow for comparison between these sites. In the ten multicenter studies included in our sample (1, 7, 12, 14, 16, 19, 21, 26, 31, and 32 in *Appendix Figure 2*), no trend in total number of findings across sites could be identified.

Cost of monitoring

Data on monitoring costs were only available since May 2012, when an electronic enterprise resource management system was implemented at the CTU. Overall, cost data was available for 33 out of 43 monitored projects. For these projects, we documented a total of 4320 working hours for 2401 monitoring findings. With an hourly salary rate of US$92.5 this translated into total personnel costs of US$ 399'280 and an average per findings cost of US$166. In this estimate, however, the endpoint related findings that were resolved on-site are not considered, leading to potential overestimation of costs per finding.

The Monitor's perspective

In addition to evaluating the characteristics of findings of our on-site visits, we aimed to understand the practical experience of monitors involved, challenges they face during monitoring, their perspectives on the risk-based approach, and suggestions for improvement.

Three monitors we interviewed had been working as clinical trial monitors for two to 21 years, two of them mainly in the academic setting whereas the third one mainly in the pharmaceutical industry environment. With the introduction of risk-based monitoring at the CTU in 2012, all three participants started to monitor according to the above described standard procedure. Below we describe three main themes from these interviews and provide select quotations for each.

Factors influencing risk-based monitoring findings

All three monitors expressed a generally positive attitude towards the concept of upfront risk evaluation, which allows assessment of critical factors in the study design or practical challenges that the study team might face while implementing the trial as described by one of our monitors.

"The positive effect clearly is that you think more about the study itself. If you take the effort to classify the study by risk factors you can actually eliminate many things upfront. Because of that evaluation, I know what to set value on when I open a site". (Monitor III)

The factors that the monitors in general deemed crucial for low numbers of findings in trials were professional, trained and motivated study personnel, together with a robust study design and rigorous planning of a study. These factors were also attributed to support participant recruitment into the trials and eventual success of the trial. Monitor III argued that indicators related to the study site were underrepresented in the ADAMON risk evaluation, in spite of the fact that they have significant influence on the way trials are conducted. Other

factors that the monitors believed to contribute positively to overall trial quality and success were the quality of the study protocol, the early involvement of monitor and study nurses in protocol development, training and experience of all personnel involved, planning of finances and infrastructure before trial begins, available resources, trial coordination and management, assignment of clear responsibilities, well planned recruitment, and clear and transparent communication among all stakeholders involved as elaborated in the quote below.

"But it all depends on the experience of staff on-site, if they have lots of experience with studies, they know how to do it. But don't forget that staff changes so often at the site, you never get the same people from the start until the end of a study. The new ones, how will they be trained? We as monitors only hear about it half a year later and if you only visit them once a year, you hear that they have changed the recruiting physician and that patients have been informed wrongly for three quarters of a year." (Monitor III)

Monitoring process and challenges faced

With respect to what the current risk-based approach is able to cover on-site, all monitors came up with two distinct topics, namely patient safety and rights, and data quality. Monitor I and II felt that minimum aspects of patient safety and rights, incl. informed consent forms and inclusion/exclusion criteria, but also "crucial" data points such as the primary endpoint, were mostly covered by their on-site visits. Two monitors did not see any issues related to patient safety or their rights with the current approach as described below.

"It depends on the monitoring plan; usually we look at 100% of the Informed Consent Forms, unless there are too many patients such as in cohort studies. We always look at the inclusion and exclusions criteria. Endpoints are to be discussed and defined with the principal investigator. Depending on the budget available, we might also look at the Trial Master File, and then I am done in no time" (Monitor I)

"I hope that I cover safety aspects with my monitoring. Actually I don't see any issues with it. There is no monitoring plan without the safety aspect, usually it is 100% covered. (…) Depending on what you find, you adapt it (the monitoring plan). If you find critical issues, for example a Serious Adverse Event that was not documented, you tell the team (on-site) to look at the other patients' data and check whether they were correct." (Monitor II)

However, Monitor III expressed concerns about rather infrequent on-site visits with long time gaps in between during which there was a clear risk of missing patient safety or patient rights

aspects in particular. He also feared that these findings would then get resolved only during the next visit which might be after six months.

"I question whether this approach is compatible with GCP. According to GCP you put patients first, and then the scientific question. There you also include data protection and the ICF (informed consent form). You don't really respect patient rights if, for example, a wrong ICF version was signed and I only notice after half a year, just because the study is a low risk study according to the evaluation. I don't think it affects the scientific validity, but more the patient safety and rights aspects". (Monitor III)

With respect to data quality, all monitors mentioned "systematic data errors", i.e. errors that are not produced by chance. They were concerned to miss systematic errors with the risk-based approach. Monitor I sometimes preferred to monitor more frequently in order to identify systematic data errors as and when they occurred. Monitor I and II would like to rather cover 100% source data for fewer patients than single puzzle pieces of several patients to be able to pick up systematic errors as described in quote below.
"One is for sure, if you don't see that mistake at the beginning, it's going to repeat with the next patients. So that's a systematic error then, and of course data quality suffers! " (…) "If I knew that every three month there is a monitor at your doorstep who wants to critically look at your data, then I am of course required to get my stuff done in time and…a bit more accurate as if, you know, I know that there is anyways no one looking at my data. Then you get the running around after the data at the end of the year". (Monitor III)

Monitor II further questioned low importance given to "less important data points" (e.g. lab values not specified as outcome variables). These "less important data points" are not considered in the risk-based monitoring plan according to ADAMON and therefore not checked by the monitors

"Often, we do not look at lab values because they are not seen as risky values, maybe only 10% is seen as crucial for the primary endpoint. But then you ask yourself whether you wouldn't miss transcription errors if you don't look at these values at all. (…) There are not only systematic errors that you don't see but also those that appear everywhere and they're even more difficult to detect. And it also depends on the format in which you collect data, if it is on paper or not. The CRF (case report form) heavily influences the number of mistakes that are made, and you don't look at all of these with the risk-based approach. (…) If you, for example, look at one patient 100% in a study and not the others, you of course detect

systematic errors and point to that and make them look at the other patients in their documentation as well." (Monitor II)

All monitors believed that data quality would improve with increased frequency of monitoring. Monitor I and III disagreed with the current guidelines that in some low-risk cases, an initiation visit is not necessary. They would rather leave out the close-out visit but always perform an initiation visit to train the personnel on crucial GCP and study-related aspects.

"I assume that in an inspection, you know, if they were to look at 100% of the data after my monitoring, they would find errors even in my monitored data. But I am sure you see exactly what documents were monitored and which weren't. (…) but I believe that data quality would for sure be better with more monitoring, there I am 100% sure." (Monitor I)

"Also the change in personnel has an influence, if you don't take care of the training of new employees, and you don't involve them. The situation on-site is not reflected adequately in ADAMON, it's more sort of a weak factor, that if you're there anyways, you monitor longer, but it doesn't influence the frequency of monitoring. Changes and structures on-site should have a stronger influence on the frequency of visits, in my opinion. "(Monitor III)

Monitor III further questioned whether in reality the risk-based approach proves cost-effective if many errors were missed in between visits due to the low frequency of visits and when amendments are needed to correct those.

"It is not clear whether it is cheaper if the monitor visits less frequently and everything on-site goes downhill or if it wasn't better if the monitor had visited once or twice more, you know. To make crappy data better again is also not cheap, right? Even if you adapt all ADAMON criteria, we should get away from only seeing cost savings in it. "(Monitor III)

When monitors were further asked about why the frequency of monitoring visits could not be adapted in cases where needed, all mentioned the difficult financial environment in which academic trials are conducted. In their experience, most often, funding limitations were the main factor for restrictions in the amount and frequency of monitoring visits that could be performed, rather than the actual risk categorization.
"…just because the budget doesn't allow, you know, I would like to monitor more frequently or follow the risk classification more strictly, but you can't, because of the financial limitations." (Monitor I)

Role of monitors and future perspectives

We explored further monitors' views on ways to improve monitoring process and to make it cost effective. All monitors discussed few possibilities especially in defining their role as monitors and the way they are perceived by the study teams. They would like to be seen as trustworthy partners who assist in ensuring trial success and quality instead of being mere "controllers". They hoped that principal investigators would be more familiar with their role and study teams will not see them as a "necessary evil" who consume significant part of the study budget and with whom they have to deal with, but rather as supporting partners who assist in achieving the study goals and ensure study quality. They see themselves as critical examiners, a fresh pair of eyes, who provide constructive feedback to the study team, as trusted supporters, facilitators of communication across sites, and motivators.

"I would wish for more acceptance of monitoring, they all think it is just a necessary evil. That you don't do it because it's important and helps data quality, but only because the regulators and authorities want it from you. I am sure, often, monitoring reports are not even read. There should be more trust that we help data quality and therefore, also help the answer to the study question". (Monitor II)

"I always tried to make them feel like I am their partner, and not some teacher or something…I don't want to point the finger at them, but build a trustworthy relationship so that people at the site know they can trust me. So that they know they can call me if there is a problem. (…) The aim is to make them understand that we are partners and try to help to get to a good result that all of us do a good job, and that patients are safe and their rights are protected". (Monitor III)

Discussion

As far as we are aware, this is the first mixed methods study to retrospectively investigate the outcome of on-site risk-based monitoring according to the ADAMON framework with regards to patient rights, safety and data quality in a sample of 43 interventional and observational studies in the academic setting. We identified a proportionate amount of patient rights and administrative findings, while findings concerning patient safety were rare, resulting in costs per documented finding of US$166. While administrative findings naturally predominantly occurred at the beginning of the studies (e.g. at initiation visit), patient rights findings developed proportionately with the proceeding enrolment of patients in the study.

Based on our study sample we found in exploratory investigation factors such as observational study design, industry sponsorship, a high risk classification, and the inclusion of a vulnerable study population to be associated with fewer findings during on-site visits. Surprisingly, a high risk category per se and the inclusion of vulnerable study populations, which come with an increased frequency of on-site monitoring visits according to ADAMON, do not cause a larger number of total findings. Counterintuitively, high risk studies therefore seem to be at lower risk for poor quality, probably due to the more closely monitored regulatory and legal environment which supports a well-planned set up of the study. It is therefore questionable whether the current monitoring scheme according to pre-set risk factors will be effective, both for quality and cost of trials, unless we learn from these findings. An alternative approach could, for example, be “experience-based” in the sense that monitoring frequency and extent are continuously adapted after on-site visits depending on the findings that have occurred. This would certainly allow for more flexible monitoring strategies when and where on-site visits are actually needed in line with current “quality-by-design” initiatives [23-25]. Practically however, changes in contracts are often difficult after monitoring plans have been written and budgets have been allocated.

Further, our results are in line with a systematic review of risk-based monitoring tools which states that both ADAMON as well as the SCTO guideline do not assess all 12 fundamental risk indicators as described in the recently published risk indicator taxonomy for supervision of clinical trials on medicinal products [26, 27]. While ADAMON lacks indicators on professionalism, reputation, and level of experience of investigator, clinical trial site, and sponsor, the SCTO guideline provides indicators assessing the level of experience at least to some extent [27]. We show that the two of the three risk modulators that the CTU has used to adjust monitoring extent purely based on experience with previous studies, i.e. the absence of an electronic data capture system and the lack of experience of a site, are clearly associated with a higher number of findings. This was exemplified by the two observational

outlier studies. These factors should therefore be considered not only in the modulation of the extent of monitoring, but also influence the frequency of on-sites visits.

Our qualitative enquiry highlighted that the involved monitors understand the positive aspects of a risk-based approach in the resource constrained academic setting, but fear to miss systematic errors or even patient right violations due to the low frequency of visits or the lack of a requirement for initiation visits in low risk studies. They stressed the importance of well trained, motivated and experienced trial personnel, i.e. the investigator, study nurses, and related site staff, for overall trial quality and success. They further exemplified that these human factors should play a larger role in the risk evaluation, and that ADAMON does not cover these aspects adequately. The additional factors that were mentioned to be crucial for trial quality and success predominantly covered the design of the study, including how well the practical aspects of the study are planned, and factors related to the functioning of the site, such as training and experience of personnel, planning of finances and infrastructure, resources, and recruitment, trial coordination and management, assignment of clear responsibilities and transparent communication. While some of these aspects covered by ADAMON, none of them has an influence on the final monitoring risk category. We therefore encourage to put more emphasis on site-related and personnel-related risk factors in the risk evaluation of any studies, both industry- and investigator initiated, in addition to any framework used.

The initial concept of risk-based monitoring aimed at optimizing the use of scarce resources while assuring patient rights and safety as well as data quality in accordance with the GCP guideline. Several years later, cornerstones of the risk-based monitoring concept, such as ADAMON and OPTIMON [8, 20], are still under evaluation and evidence on the effectiveness and cost savings of this approach in different host organizations, sponsors, settings, or trial designs remains relatively sparse [28, 29]. There is, however, emerging consensus that 100% SDV and dual entry procedures are time and cost inefficient in detecting data discrepancies [17, 30, 31], that some types of errors in a clinical trial are more important than others [6], and that a site monitoring approach tailored to the risk of the trial can be supported in order to detect critical issues [7, 8, 32]. With our approach, we mainly identified patient rights and administrative findings, for which we were not able to retrospectively judge how critical they were. Our monitors, however, clearly stated the fear of missing systematic data errors and even critical issues concerning patient rights during the on-site visits, which are restricted in frequency and extent by academic financial constraints, or a low risk category. We thus see a need for complementary quality assurance measures for systematic data errors that can be performed off-site, such as central data verification.

Recently, combinations of central data verification (e.g. patient rights and safety) and on-site monitoring strategies have been applied to improve the efficiency of risk-based procedures [5, 7, 33]. Compared to other tools, ADAMON does only provide vague guidance on the nature and extent of centralized monitoring, while the adapted form by the Swiss Clinical Trial Organization recommends that protocol compliance could be monitored centrally for low-risk trials, as described in a recent systematic review by Hurley et al [27]. According to the FDA's recommendations on risk-based monitoring for industry and the current integrated addendum to ICH-GCP, centralized monitoring processes could provide additional capabilities to on-site monitoring in the academic setting, thereby dispelling our monitor's doubts on missing systematic data errors [3, 9]. Statistical monitoring methods are an area of active research and have been suggested to "help improve the effectiveness of on-site monitoring by prioritizing site visits and by guiding site visits with central statistical data checks" [33] and were shown to identify the great majority of on-site monitoring findings [7]. In line with our results, Tudur Smith et al. have recently described such an approach in non-commercial studies, allowing for "triggered", rather than predominantly "routine" on-site visits [5]. However, empirical data on its effectiveness and the costs, advantages and disadvantages of alternative methods are still missing. With central monitoring strategies allowing for efficient data quality checks, monitors would then also be more flexible to transform their roles from "controllers" towards "partners", as they had mentioned in the interviews, and contribute to overall clinical research quality rather than mere GCP-conformance.

We are aware of a number of limiting factors in our analysis as follows. First of all, the retrospective design of our study did not allow standardization of extracted data across studies and monitors. Our sample was heterogeneous in terms of study type, study design, intended sample size and the risk categorization. It may further not be entirely representative in type and size of all studies conducted at out institution, as we predominantly monitor investigator-initiated studies. However, we aimed to minimize selection bias by including all trials monitored by the CTU within a given time period. Second, due to the small sample size and large heterogeneity of studies, we did not perform multivariable analysis but summarized absolute numbers of findings descriptively. Third, these studies were monitored by six different monitors with varying degree of experience and over a period of two years. In spite of a standardized procedure, each monitor has his or her own personal monitoring style, different level of attention to detail and strive for perfection. These factors could have influenced their interactions with the study team, generation and documentation of monitoring related findings. Fourth, we could only interview three of these six monitors who continue to work at our institution. We are fully aware that their experience and perspective cannot be

generalized but their perspective is critical in understanding the challenges in effective monitoring and ways to improve monitoring process. Furthermore, perspectives of additional stakeholders involved in risk-based monitoring (e.g. trial project leaders or principal investigators) should be considered in the future. Finally, we want to discuss the subjective and flexible nature of risk classification (e.g. risk compared to standard treatment, judging the experience of staff at a study site and adjustments to available budgets) and diversity in monitoring style of different monitors and their documentation practice. This could have contributed to an unknown number of findings that were resolved directly on-site without documentation and hence out of scope of current analysis. In addition, we were unable to calculate number of findings in relation to number of patients recruited due to inconsistent documentation of number of patients monitored during each monitoring visit.

However, by publishing our experiences, we are supporting an ancillary recommendation made by the Clinical Trials Transformation Initiative (CTTI) project on effective and efficient monitoring to “share knowledge and experiences, so that best practices may be established” [34]. The hurdles that we experienced in the generation of urgently needed evidence on effectiveness of monitoring strategies have provided food for thought as follows: Our study illustrates that while the very concept of the current risk-based approach allows for flexibility in tailoring monitoring to the requirements of a specific study, it may also be prone to ambiguity. The lack of a "one-fits-all" model may lead stakeholders to be at loss to define a relevant trial-specific monitoring strategy. We realized that the combination of flexibility in designing the monitoring approach and the subjectivity and individual preferences of monitors in documenting findings add to the complexity of analyzing the effectiveness of risk-based approaches. For instance, the change of monitor in a project was often associated with a specific findings pattern (e.g. an increase in findings), which may be explained by preferences regarding the individual documentation style, the level of detail monitored, or the monitor’s experience. This haziness including the inter-human and inter-institutional variability should receive more recognition in the field when investigating the effectiveness of risk-based approaches in the academic setting.

Finally, we perceive that all of the efforts invested so far focus on optimizing the cost-effectiveness of current strategies, i.e. assuring patient’s rights and safety as well as data quality in accordance with the requirements of ICH-GCP. Interestingly, none of the efforts has so far questioned the real impact of current monitoring activities in increasing the overall quality of academic clinical studies. In line with other authors we believe that the current detection of non-critical findings adds only little to overall study quality, while consuming significant resources that could be spent in areas known to be critical for successful trial

conduct [17]. For instance, empirical data show that academic clinical research in particularly suffers from major hurdles in the recruitment of patients, with recruitment failure being the main reason for early discontinuation of trials [35-39]. According to a report by the Tufts Center for the Study of Drug Development, about 50 % of sites fail to reach the planned recruitment targets and more than 95 % of clinical trials do not end on time and on budget as planned in the first place. Ninety percent of the studies meet their recruitment goals, but at the expense of mostly twice as much time as originally planned [40]. Often, this is due to too many avoidable protocol amendments, with a first amendment implemented even before the very first patient has been enrolled [41]. Almost half of the protocol amendments are considered “somewhat” or “completely” avoidable [40].The European Clinical Research Infrastructure Network (ECRIN) hence provides a broad definition of monitoring as activities which must be “understood as all onsite and central activities dealing with checks of data and procedures as well as with the overall surveillance and stimulation of the trial progress” [42]. The interviewed monitors have supported such a holistic approach starting with their involvement in protocol development and processes based on trial procedures rather than data points per se, supporting overall trial completion with conclusive results. In accordance with ICH-GCP E6 R(2), we envision the future monitor to be an on-site partner to the study team, supported by centralized data checks adaptable to the risk of a trial, considering the experience of and the management at the site itself [3].

Conclusion

In conclusion, we show that the factors which mostly increased the risk for on-site monitoring findings are underrepresented in the current ADAMON scheme, but have partly been considered by our monitors based on their professional experience. We believe that the “human factor” has been underestimated in the evaluation of risk-based approaches so far, and should receive more recognition in the future. In line with recent developments, our risk-based on-site monitoring should be complemented by centralized data checks in the future, allowing monitors to transform their role towards partners for overall trial quality, and success. However, evidence on the methodology and the (cost-) effectiveness of different combinations of the two approaches is still sparse for the academic setting. Future research should therefore address urgently needed strategies for efficient and effective monitoring, based on the current knowledge on risk factors in the academic setting.

Declarations

Ethics approval and consent to participate: As interviews did not include health-related data and were therefore not within the scope of the applicable Swiss Human Research Act (HRA, Art.1), we did not require formal ethical approval but obtained a Declaration of no obligation (Req-2016-00744) from the Ethics Committee northwest/central Switzerland (EKNZ). Consent from monitors who were interviewed was obtained before study start.

Acknowledgements

We would like to thank the monitors for the interviews.

References

[1] Castle PE, Cuzick J, Stoler MH, Wright TC, Jr., Reid JL, Dockter J, et al. Detection of Human Papillomavirus 16, 18, and 45 in Women With ASC-US Cytology and the Risk of Cervical Precancer: Results From the CLEAR HPV Study. Am J Clin Pathol. 2015;143:160-7.
[2] ICH Harmonised Tripartite Guideline: Guideline for Good Clinical Practice. 2001.
[3] International conference on harmonisation of technical requirements for registration of pharmaceuticals for human use, Intergrated Addendum to ICH E6(R1): Guideline for Good Clinical Practice E6R(2)

[4] Tudur Smith C, Stocken DD, Dunn J, Cox T, Ghaneh P, Cunningham D, et al. The value of source data verification in a cancer clinical trial. PLoS ONE. 2012;7:e51623.
[5] Tudur Smith C, Williamson P, Jones A, Smyth A, Hewer SL, Gamble C. Risk-proportionate clinical trial monitoring: an example approach from a non-commercial trials unit. Trials. 2014;15:127.
[6] Baigent C, Harrell FE, Buyse M, Emberson JR, Altman DG. Ensuring trial validity by data quality assurance and diversification of monitoring methods. Clin Trials. 2008;5:49-55.
[7] Bakobaki JM, Rauchenberger M, Joffe N, McCormack S, Stenning S, Meredith S. The potential for central monitoring techniques to replace on-site monitoring: findings from an international multi-centre clinical trial. Clin Trials. 2012;9:257-64.
[8] Journot V, Pignon JP, Gaultier C, Daurat V, Bouxin-Metro A, Giraudeau B, et al. Validation of a risk-assessment scale and a risk-adapted monitoring plan for academic clinical research studies--the Pre-Optimon study. Contemp Clin Trials. 2011;32:16-24.
[9] Risk based approach to monitoring: Guidance for Industry. US Food And Drug Administration FDA; 2013.
[10] European Commission, Risk Proportionate Approaches in Clinical Trials-Recommendations of the Expert Group on Clinical Trials for the Implementation of Regulation (EU) no 536/2014 on Clinical Trials on Medicinal Products for Human Use, 2014.
[11] Brosteanu O, Houben P, Ihrig K, Ohmann C, Paulus U, Pfistner B, et al. Risk analysis and risk adapted on-site monitoring in noncommercial clinical trials. Clin Trials. 2009;6:585-96.
[12] European Medicines Agency: Reflection paper on risk based quality management in clinical trials. 2014.
[13] Organisation for Economic Co-operation and Development (OECD), OECD recommendation on the governance of clinical trials, 2013.
[14] Funning S GA, Eriksson K, Kettis-Linblad Å. Quality assurance within the scope of Good Clinical Practice (GCP) – what is the cost of GCP-related activities? A survey within the Swedish Association of the Pharmaceutical Industry (LIF)'s members. The Quality Assurance Journal. 2009;12:3-7.
[15] Vantongelen K, Steward W, Blackledge G, Verweij J, Van Oosterom A. EORTC joint ventures in quality control: treatment-related variables and data acquisition in chemotherapy trials. Eur J Cancer. 1991;27:201-7.
[16] Rosa C, Campbell A, Kleppinger C, Sampson R, Tyson C, Mamay-Gentilin S. Quality assurance of research protocols conducted in the community: the National Institute on Drug Abuse Clinical Trials Network experience. Clin Trials. 2009;6:151-61.
[17] Pronker E, Geerts BF, Cohen A, Pieterse H. Improving the quality of drug research or simply increasing its cost? An evidence-based study of the cost for data monitoring in clinical trials. Br J Clin Pharmacol. 2011;71:467-70.
[18] Olsen R, Bihlet AR, Kalakou F, Andersen JR. The impact of clinical trial monitoring approaches on data integrity and cost-a review of current literature. Eur J Clin Pharmacol. 2016;72:399-412.
[19] Bakobaki J, Joffe N, Burdett S, Tierney J, Meredith S, Stenning S. A systematic search for reports of site monitoring technique comparisons in clinical trials. Clin. 2012;9:777-80.
[20] Brosteanu O, Houben P, Ihrig K, Ohmann C, Paulus U, Pfistner B, et al. Risk analysis and risk adapted on-site monitoring in noncommercial clinical trials. Clin. 2009;6:585-96.

[21] Guidelines for Good Operational Practice V2.0. In: Organisation SCT, editor.2014.
[22] Brassil D, Kost RG, Dowd KA, Hurley AM, Rainer TL, Coller BS. The Rockefeller University Navigation Program: a structured multidisciplinary protocol development and educational program to advance translational research. Clin Transl Sci. 2014;7:12-9.
[23] Meeker-O'Connell A, Glessner C, Behm M, Mulinde J, Roach N, Sweeney F, et al. Enhancing clinical evidence by proactively building quality into clinical trials. Clin Trials. 2016;13:439-44.
[24] Landray MJ, Grandinetti C, Kramer JM, Morrison BW, Ball L, Sherman RE. Clinical Trials: Rethinking How We Ensure Quality. Drug Inf J. 2012;46:657-60.
[25] Reith C, Landray M, Devereaux PJ, Bosch J, Granger CB, Baigent C, et al. Randomized clinical trials--removing unnecessary obstacles. N Engl J Med. 2013;369:1061-5.
[26] Jongen PMJM, van den Bogert CA, van de Laar CWE, Notenboom K, Hille ETM, Hegger I. Risk indicator taxonomy for supervision of clinical trials on medicinal products. Current Medical Research and Opinion. 2016;32:1269-76.
[27] Hurley C, Shiely F, Power J, Clarke M, Eustace JA, Flanagan E, et al. Risk based monitoring (RBM) tools for clinical trials: A systematic review. Contemp Clin Trials. 2016;51:15-27.
[28] Macefield RC, Beswick AD, Blazeby JM, Lane JA. A systematic review of on-site monitoring methods for health-care randomised controlled trials. Clin Trials. 2013;10:104-24.
[29] Morrison BW, Cochran CJ, White JG, Harley J, Kleppinger CF, Liu A, et al. Monitoring the quality of conduct of clinical trials: a survey of current practices. Clin Trials. 2011;8:342-9.
[30] Andersen JR, Byrjalsen I, Bihlet A, Kalakou F, Hoeck HC, Hansen G, et al. Impact of source data verification on data quality in clinical trials: an empirical post hoc analysis of three phase 3 randomized clinical trials. Br J Clin Pharmacol. 2015;79:660-8.
[31] Lindblad AS, Manukyan Z, Purohit-Sheth T, Gensler G, Okwesili P, Meeker-O'Connell A, et al. Central site monitoring: results from a test of accuracy in identifying trials and sites failing Food and Drug Administration inspection. Clin Trials. 2014;11:205-17.
[32] Blessing JA, Bialy SA, Whitney CW, Stonebraker BL, Stehman FB. Gynecologic Oncology Group quality assurance audits: analysis and initiatives for improvement. Clinical Trials. 2010;7:390-9.
[33] Venet D, Doffagne E, Burzykowski T, Beckers F, Tellier Y, Genevois-Marlin E, et al. A statistical approach to central monitoring of data quality in clinical trials. Clin Trials. 2012;9:705-13.
[34] Effective and Efficient Monitoring as a Component of Quality.
[35] Kasenda B, von Elm E, You J, Blumle A, Tomonaga Y, Saccilotto R, et al. Prevalence, characteristics, and publication of discontinued randomized trials. Jama. 2014;311:1045-51.
[36] Kasenda B, von Elm EB, You J, Blumle A, Tomonaga Y, Saccilotto R, et al. Learning from failure--rationale and design for a study about discontinuation of randomized trials (DISCO study). BMC Med Res Methodol. 2012;12:131.
[37] Rosenthal R, Kasenda B, Dell-Kuster S, von Elm E, You J, Blumle A, et al. Completion and Publication Rates of Randomized Controlled Trials in Surgery: An Empirical Study. Ann Surg. 2015;262:68-73.
[38] Schandelmaier S, von Elm E, You JJ, Blumle A, Tomonaga Y, Lamontagne F, et al. Premature Discontinuation of Randomized Trials in Critical and Emergency Care: A Retrospective Cohort Study. Crit Care Med. 2016;44:130-7.
[39] Stegert M, Kasenda B, von Elm E, You JJ, Blumle A, Tomonaga Y, et al. An analysis of protocols and publications suggested that most discontinuations of clinical trials were not based on preplanned interim analyses or stopping rules. J Clin Epidemiol. 2016;69:152-60.
[40] Tufts Center for the Study of Drug Development Impact Report "New Research from Tufts Characterizes Effectiveness and Variability of Patient Recruitment and Retention Practices". 2013.
[41] Getz KA, Zuckerman R, Cropp AB, Hindle AL, Krauss R, Kaitin KI. Measuring the incidence, causes, and repercussions of protocol amendments. Drug Inf J. 2011;45:265-75.
[42] ECRIN Risk-Based Monitoring Tool Box.

Supplementary Material

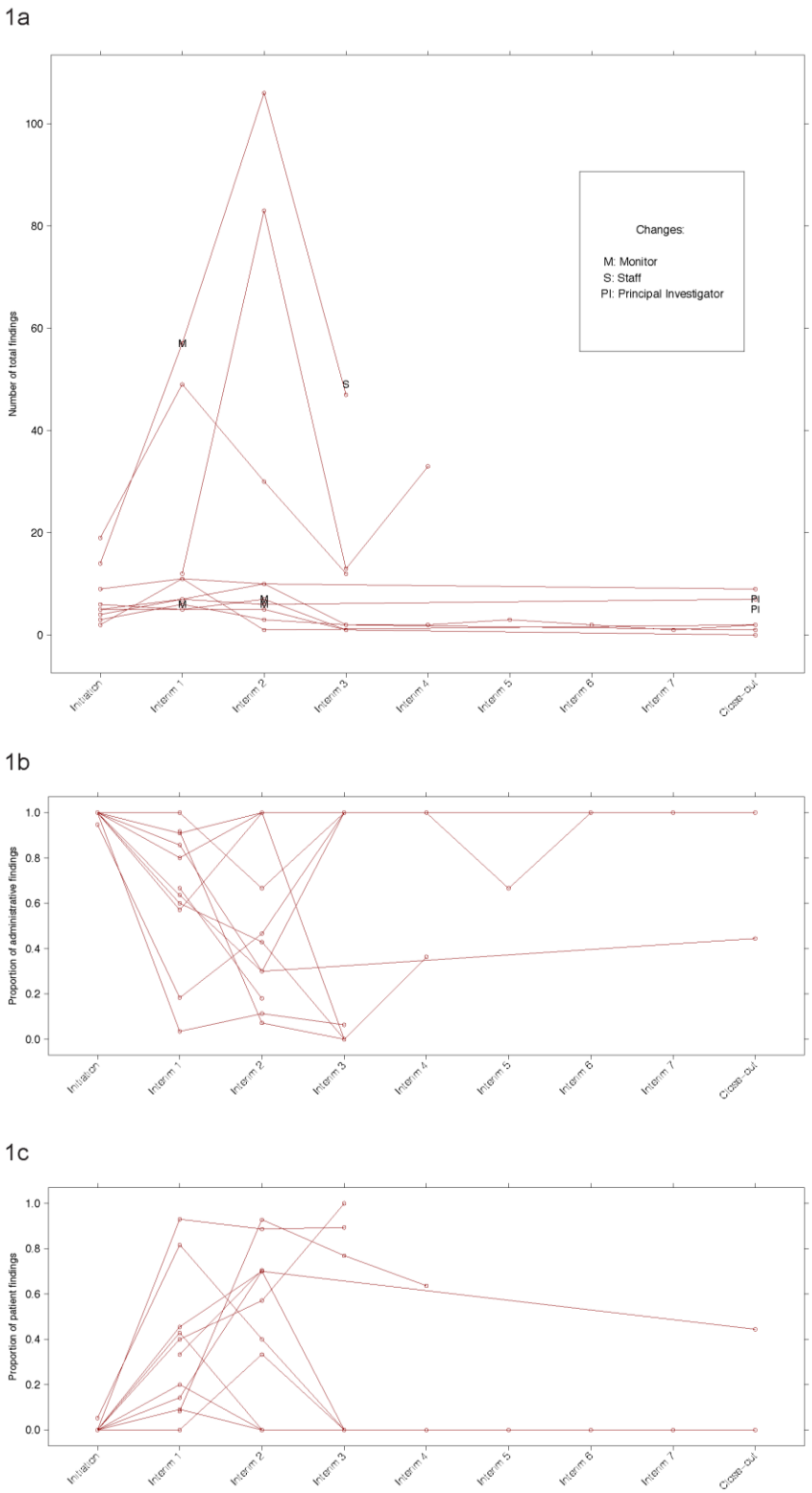


Figure 1. Total number (a) and proportion of findings (b/c) over time, per site. In (a), only studies (sites) with 3 or more monitoring visits are presented.

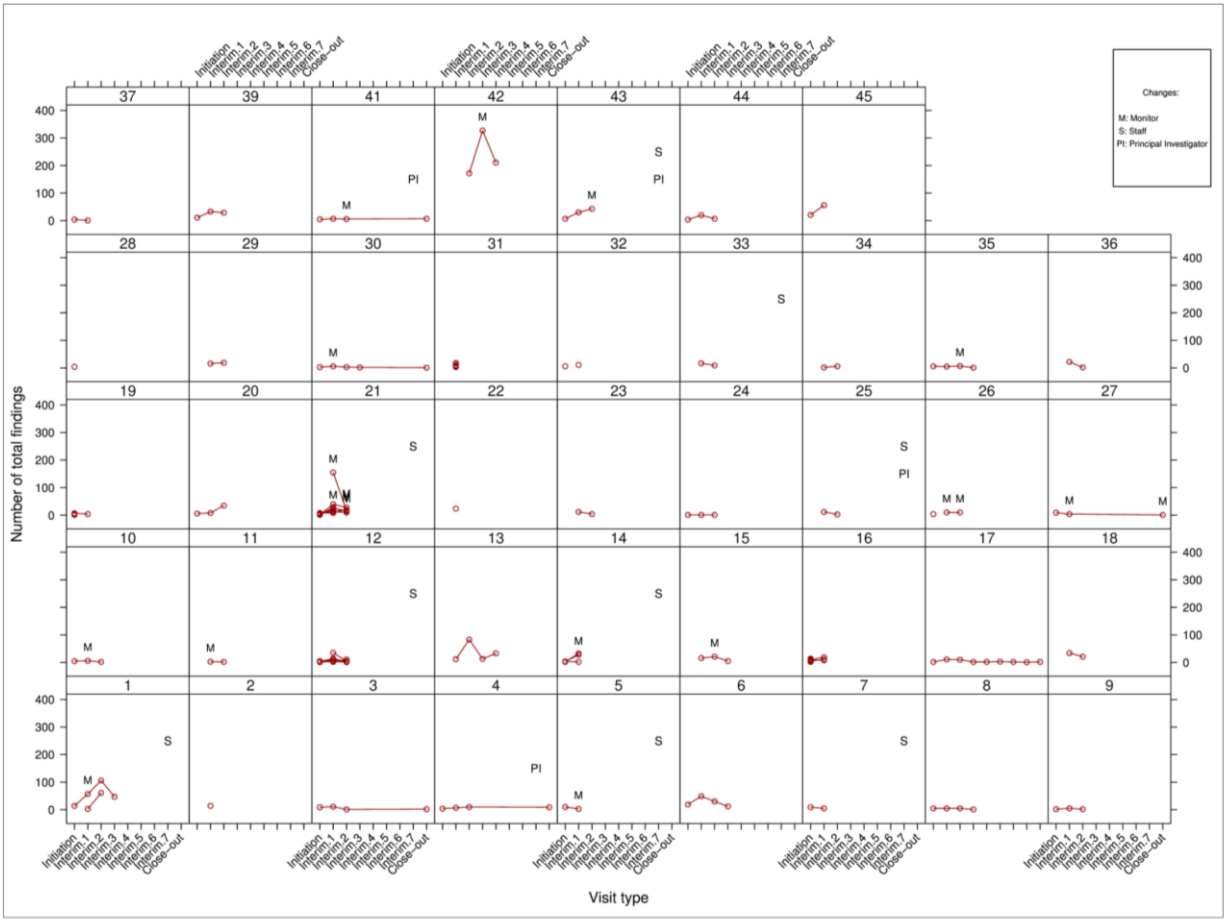


Figure 2. Total number of findings over time, by individual study and site. Circles depict monitoring visit, lines connect visits at one particular site. Circles that are not connected by lines depict monitoring visits at different sites. Number 1, 7, 12, 14, 16, 19, 21, 26, 31, and 32 are multicenter studies. If different sites are not distinguishable, the total number of findings at this particular visit was the same (superposed circles).

Manuscript VI: Validity of mobile electronic data capture in clinical studies: A pilot study in a pediatric population

Belinda von Niederhäusern[1#], MSc, MSc; Ramon Saccilotto[2], MD, MAS; Sabine Schädelin[1], MSc; Victoria Ziesenitz[3], MD, MSc; Pascal Benkert[1], PhD; Marie-Luise Decker[3], MD; Anya Hammann[1], PhD; Julia Bielicki[3,4], MD; Marc Pfister[3], MD; Christiane Pauli-Magnus[1,2], MD

[1] Clinical Trial Unit, Department of Clinical Research, University and University Hospital of Basel, 4031 Basel, Switzerland (belinda.vonniederhausern@usb.ch, sabine.schaedelin@usb.ch, pascal.benkert@usb.ch, anya.hammann@usb.ch)

[2] Department of Clinical Research, University and University Hospital of Basel, 4031 Basel, Switzerland (ramon.saccilotto@usb.ch, Christiane.pauli-magnus@usb.ch)

[3] Division of Paediatric Pharmacology & Pharmacometrics, University of Basel Children's Hospital, 4031 Basel, Switzerland (Victoria.Ziesenitz@ukbb.ch, marieluise_decker@yahoo.com, Marc.Pfister@ukbb.ch)

[4] Department of Paediatric Infectious Diseases, University of Basel Children's Hospital, 4031 Basel, Switzerland (JuliaAnna.Bielicki@ukbb.ch)

Status: Under review, BMC Medical Research Methodology

Abstract

Background: Clinical studies in children are necessary yet conducting multiple visits at study centers remains challenging. The success of “care-at-home” initiatives and remote clinical trials suggests their potential to facilitate conduct of pediatric studies. This pilot aimed to study the feasibility of remotely collecting valid (i.e. complete and correct) saliva samples and clinical data utilizing mobile technology.

Methods: Single-center, prospective pilot study in children undergoing elective tonsillectomy at the University of Basel Children’s Hospital. Data on pain scores and concomitant medication and saliva samples were collected by caregivers on two to four inpatient study days and on three consecutive study days at home. A tailored mobile application developed for this study supported data collection. The primary endpoint was the proportion of complete and correct caregiver-collected data (pain scale) and saliva samples in the at-home setting. Secondary endpoints included the proportion of complete and correct saliva samples in the inpatient setting, subjective feasibility for caregivers, and study cost.

Results: A total number of 23 children were included in the study of which 17 children, median age 6.0 years (IQR 5.0, 7.4), completed the study. During the at-home phase, 71.9% [CI= 64.4, 78.6] of all caregiver-collected pain assessments and 53.9% [CI= 44.2, 63.4] of all saliva samples were complete and correct. Overall, 64.7% [CI=58.7, 70.4] of all data collected by caregivers at home was complete and correct. The predominant reason for incorrectness of data was adherence to the timing of predefined patient actions. Participating caregivers reported high levels of satisfaction and willingness to participate in similar trials in the future. Study costs for a potential sample size of 100 patients were calculated to be 20% lower for the at-home than for a traditional in-patient study setting.

Conclusions: Mobile device supported studies conducted at home may provide a cost-effective approach to facilitate conduct of clinical studies in children. Given findings in this pilot study, data collection at home may focus on electronic data capture rather than biological sampling.

Background

High quality research relies on the collection of high quality data. Traditionally, this is done in the inpatient setting or through ambulatory visits to a study site, which can present a barrier to participation (e.g. cost, travel burden, time) and a risk to the validity of research (e.g. high loss-to-follow-up, low external validity) [1, 2]. The widespread availability of new technologies has the potential of shifting some research activities, including enrollment, managing trial activity, reporting results, and safety oversight, away from study sites. Such “remote” research may encourage the participation of a more diverse group of patients in research with improved recruitment rates and at lower costs than those of conventional trials [3-5]. A combinatory approach including direct interactions with the study team may allow remote data collection to be optimally leveraged [6, 7], e.g. by addressing challenges around data quality and retention [8-11].

In pediatric care, empirical evidence on the optimal dosing and action of routinely used medicine remains limited [12-14]. The relationship between drug exposure and its effects are often different in children compared to adults. For this reason a simple extrapolation of pediatric dosing based on adult data can put children at increased risk of adverse events and therapeutic failures [15-17]. Therefore, innovative clinical study designs in pediatrics are urgently needed. Currently, major challenges of designing and conducting clinical trials in children, include (i) small sample sizes of pediatric studies, (ii) increased study complexity due to multiple age groups, (iii) integration of research in daily activities of the whole family affecting parental time of work and supervision of other children, and (iv) child absence of routine activities. Together with the burdens of travel and frequent site visits, these limits are associated with low recruitment and high dropout rates [18, 19].

Recent „care-at-home“-initiatives indicate that mobile or remote approaches in clinical research with children have the potential to increase patient and caregiver satisfaction without increasing privately borne costs [20]. In addition, today’s parents and their children are technology savvy and frequent users of mobile devices, suggesting their high potential as future candidates in remote clinical trials. While the methodology is still in its infancy, increasing interest and support from regulators, sponsors, and patients will likely propel remote trials forward in the near future [21]. Thus, further investigation into the methodology of studies that use a combinatory approach is warranted. In this pilot study, we aimed to investigate the feasibility of

remotely collecting valid (i.e. complete and correct) clinical data and saliva samples in a pediatric population utilizing mobile technologies. In addition, we assessed the general acceptance, reasons for non-consent, and the resulting costs of this study.

Methods

Study design

This was an investigator-initiated, single center, prospective pilot study investigating the feasibility of remotely conducting clinical studies with parents/legal representatives ("caregivers") and children. We developed a custom mobile application ("app") allowing patients and their caregivers to participate in the study remotely after an initial training session at our institution (details of application development may be found in Additional file 1). We elected pain management after tonsillectomy as a model due to the frequency of the surgical procedure in children younger than 15 years [22] requiring standardized analgesic therapy and the potential to remotely assess pain levels by caregivers using a validated scale (The Childhood Discomfort and Pain Scale) [23]. In addition, remotely collect saliva samples were planned to measure acetaminophen concentrations mimicking the design of a pharmacokinetic (PK) study. In addition, the local standard of care sequence after tonsillectomy consisting of two to four days inpatient care after surgery allowed study staff to train caregivers in the use of study technologies and procedures for the at-home phase.

Total study duration for each participant and caregivers was 10 days during which data and samples were collected on 2-4 days as an inpatient and on 3 days at home. On day 8, the caregivers filled out a feasibility questionnaire. On day 10, study staff additionally contacted caregivers for feedback on the study in a follow-up telephone interview.

Participant eligibility

Children presenting for elective tonsillectomy were screened and enrolled at the University of Basel Children's Hospital from May 26 2016 until January 07 2017, during the pre-surgery anesthetics consultation. Inclusion criteria were age between 2-10 years, routine elective tonsillectomy (with or without other additional Ear, Nose, and Throat intervention), anticipated inpatient stay of a minimum of 2 days, willingness and ability of caregivers to understand and implement study procedures

in the hospital and at home, and ability of caregivers to understand, speak, and read German. Exclusion criteria were contraindications to acetaminophen administration and any reasons precluding the collection of saliva samples.

Study Procedures

Screening of eligible patients was performed at the pre-anaesthetics clinic consultation by a physician and a study nurse. After assessment of inclusion/ exclusion criteria and written informed consent, the study nurse informed caregivers about the mobile study application, provided them with an instruction manual, and supported them in the setup of their login. Caregivers had the possibility to choose between their own mobile phone (Bring Your Own Device, “BYOD”), or an iPod-Touch provided by the study team for the duration of the study.

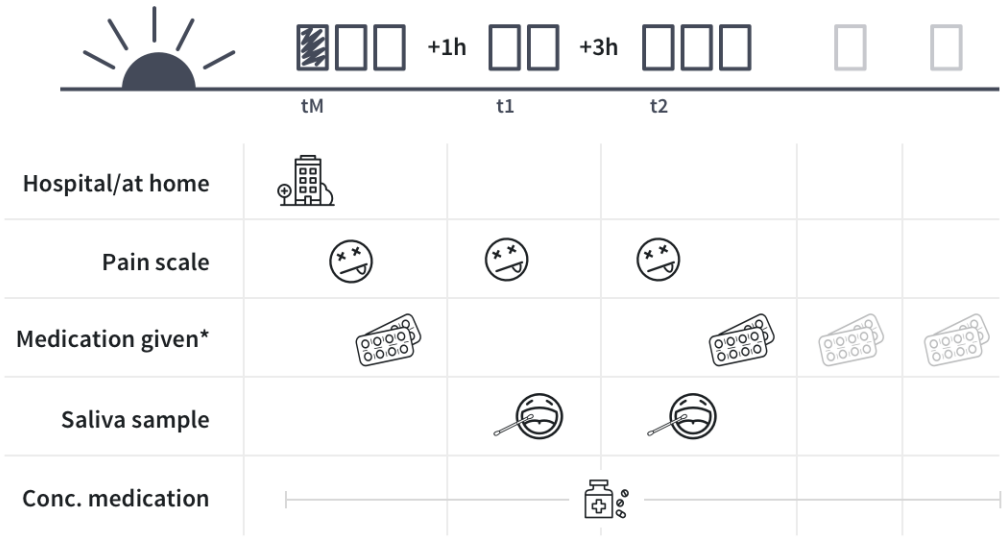


Figure 1. Daily data collection schedule.
tM: Timepoint directly after awakening of child, t1: 1 hour (+/- 15 min) after administration of first routinely scheduled dose of acetaminophen; t2: 4 hours (+2 hours) after administration of first routinely scheduled dose of acetaminophen.
* “Medication given” indicates the timepoint at which children had either received routine acetaminophen, or not (yes/no). Independent of whether medication was given or not, the app used the recorded time stamp to automatically calculate t1 and t2.

On the day of tonsillectomy, a study nurse explained the procedures, data collection and the Childhood Discomfort and Pain Scale to participating children and caregivers. After the surgeon performed tonsillectomy, postoperative pain management with acetaminophen was initiated according to the current standard of

care of the hospital [24]. The study did not interfere with routine pain management. Participating children stayed on the ward for approximately 3 days.

Daily data and sample collection was scheduled to mimic a pharmacokinetics- and dynamics (PK/PD) study. The mobile app issued automatic electronic reminders for scheduled doses as well as pain assessment/sample collection time points. Data entry was automatically time stamped. Pain assessment was repeated 3 times a day by the caregivers upon awakening of the participating children in the morning (tM) and 1 (+/- 15 min, t1) and 4 (+2 hours, t2) hours after administration of the first routinely scheduled dose of acetaminophen. Saliva samples were collected twice daily for each participant at t1 and t2. Sample codes were either scanned using the mobile application scanning function or typed in by caregivers. All concomitant medication allowed according to the guidelines of the hospital was documented throughout the study by taking a photograph of the blister or package using the mobile application. After initial supervision by a study nurse, caregivers conducted these assessments autonomously while still in the inpatient setting.

On day 3(+/- 1 day) post-surgery, participating children were discharged home. The study nurse explained how to collect and store saliva samples at home and provided labeled containers for all samples. Children staying in hospital longer than 4 days were excluded from the study. On the 3 days following discharge, caregivers collected data and samples at home following the scheme established during the inpatient stay. On study day 8, a bicycle messenger collected all saliva samples. The mobile application reminded caregivers to fill in a feasibility questionnaire before uploading data to the study server. On day 10, a study nurse conducted a follow-up telephone call for general caregiver feedback.

Statistical analyses

Primary and secondary endpoints

The primary endpoint of the study was the proportion of simultaneously *complete* and *correct* data (pain scale or saliva) in the at-home setting. A complete data set consisted of five data points (three pain scale assessments and two saliva samples) for each day and patient. Complete pain scale data were considered correct if collected within the predefined timeframes (1 hour (+/- 15 minutes) and 4 hours (+2 hours) after first medication). Complete saliva samples were considered correct if i) collected within predefined timeframe, ii) saliva volume sufficient for potential laboratory analyses, and iii) unique sample ID entered into mobile application. In the

initial analysis plan, we aimed to measure acetaminophen levels which were omitted due to technical limitations in reliably detecting acetaminophen in saliva. Secondary endpoints included the proportion of complete and correct samples in the inpatient setting (training setting before hospital release, collected as described for the out-patient setting), the subjective feasibility for caregivers as measured by an electronic questionnaire on day 8, the percentage of consenting patients, the reasons for non-consent, the legibility of photos of concomitant medication using the mobile application, and study cost.

Primary Analysis

The full analysis set consisted of all patients and caregivers who fulfilled all inclusion criteria and consented to take part. We descriptively summarized the proportion of complete and correct clinical data (pain scale) and samples (saliva) collected in the at home setting. In addition, the number of complete and correct samples per patient was modeled in a logistic regression and the 95% confidence interval (CI) was estimated based on profiled log-likelihood functions. Based on our experience with paper-based patient-reported outcomes (e.g. questionnaires), our hypothesis was that the overall completeness and correctness of electronically collected data would be above a predefined threshold of 90 %.

Secondary Analyses

The proportion of simultaneously *complete* and *correct* data (pain scale or saliva) in the inpatient setting was analyzed as described for the primary endpoint. In addition, both analyses were repeated with a secondary analysis set consisting of 15 patients who used the newest version of the mobile application. Baseline characteristics, study flow statistics, reasons for non-consent, feasibility questionnaires, and legibility of images were summarized and presented descriptively.

Missing data and drop-outs

Missing data were part of the primary endpoint (completeness not reached). Patients who were re-hospitalized within three days after discharge were considered drop-outs, and reasons were documented. All other data were assumed to be missing at random and no imputations were performed.

Cost analysis

We describe a total study cost approach factoring in app development and testing, on-site, data management, analysis staff time, study-specific materials, laboratory sample analysis, and transport costs. Cost calculations were based on salaried staff time log sheets and fixed costs for materials. Sensitivity analyses include cost for a traditional, fully on-site conducted scenario, and cost for larger samples size studies.

Results

Patient and caregiver characteristics

Of the 45 patients and their caregivers assessed for eligibility, twenty-three (51.1%) were enrolled in the study, and 17 (37.8%) completed the full study (Figure 2). Of the 15 (33.3%) caregivers who declined to participate, thirteen consented to provide a reason for non-consent, which predominantly included the perceived time burden of the study (8/13, 61.5%) (Table 1). Of 23 enrolled patients, 6 (26%) dropped out during study conduct. Reasons included one non study-related serious adverse event, patient refusal to provide saliva samples, technical issues with the mobile application, and contraindications to acetaminophen (Figure 2).

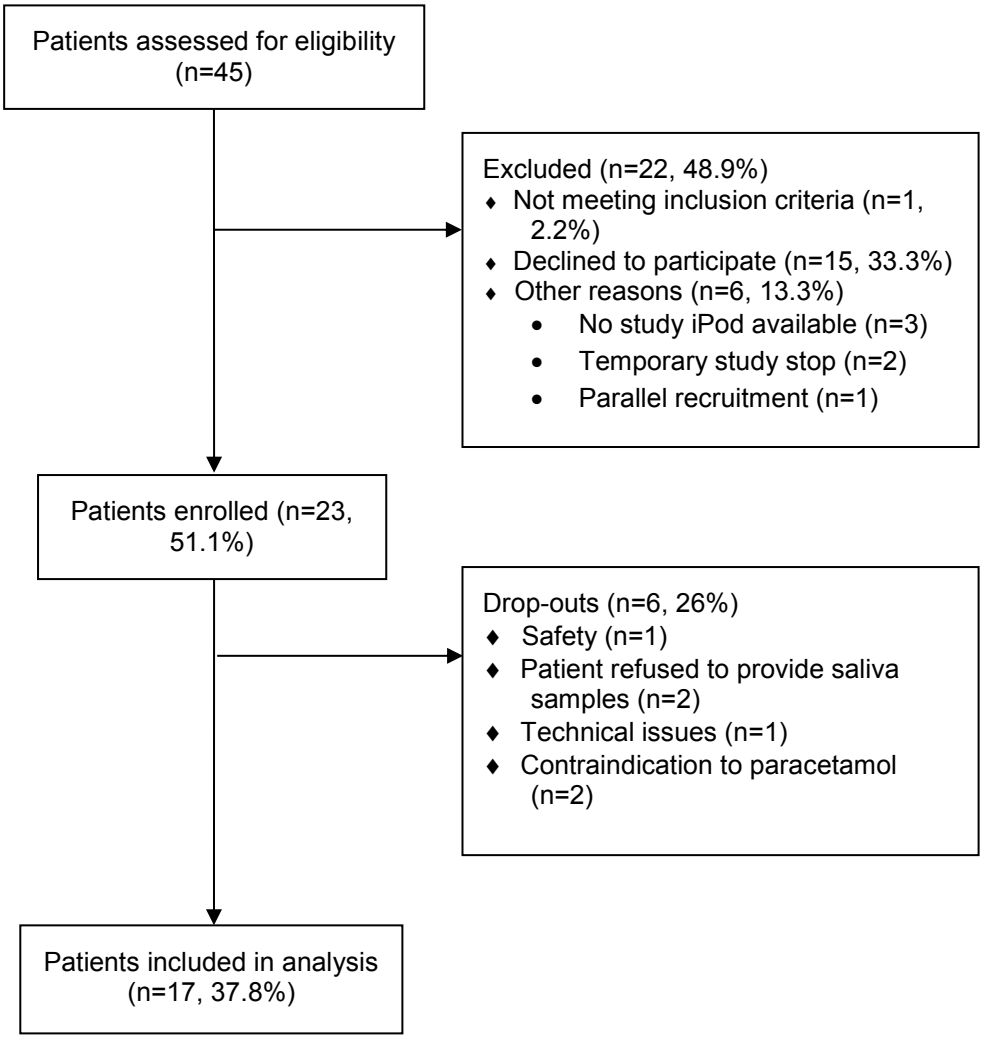


Figure 2. TOMACHI Study Flow Diagram

Table 1. Reasons for caregiver non-consent

	n (%)
Caregivers who declined to participate (n=15) and provided reason for non-consent	13 (86.7)
I do not have the time to conduct the study	8 (61.5)
I do not believe I can collect data and samples correctly	4 (30.8)
I do not want to put additional burden on my child	3 (23.1)
I did not fully understand what the study is about	2 (15.4)
I generally have doubts about clinical research	0 (0)
I would be interested to participate in such a study in the future	6 (46.2)

Baseline characteristics of the 17 patients and caregivers who completed the study are described in Table 2. Median age of patients and caregivers was 6.0 (Interquartile Range (IQR) 5.0-7.4) and 35.0 (IQR 32.0-38.0), respectively. A majority of patients (11/17, 64.7%) had one sibling, and 14/17 caregivers (82.4%) were native German speakers

Table 2. Baseline characteristics of patients and caregivers

n	17	(100)
Patient gender (n male (%))	10	(58.8)
Age (median years [IQR])	6.0	[5.0, 7.4]
Number of siblings (%)		
0	2	(11.8)
1	11	(64.7)
2	2	(11.8)
3	2	(11.8)
Caregiver age (median years [IQR])	35.0	[32.0, 38.0]
Caregiver native German speaker = yes (%)	14	(82.4)
Caregiver working at the moment = yes (%)	12	(70.6)
Caregiver occupation (ISCO) (%)		
At home/unemployed	5	(29.4)
Professional	3	(17.6)
Service and sales workers	9	(52.9)
Caregiver volume of work (median weekly % [IQR])	60.0	[40.0, 70.0]

IQR, Interquartile Range; ISCO, International Standard Classification of Occupations

Completeness and correctness of saliva sampling and pain assessments at home

In total, caregivers collected 303 pain scale assessments and 202 saliva samples. During the at-home phase, 71.9% [CI= 64.4, 78.6] of all pain assessments were complete and correct (92.2% complete, and thereof 78.0% correct) compared to 53.9% [CI= 44.2, 63.4] (77.5% complete, and thereof 69.6% correct) of all saliva samples (Table 3). Overall, 64.7% [CI=58.7, 70.4] of all data collected by caregivers at home was complete and correct.

Completeness and correctness of saliva sampling and pain assessments in the inpatient setting

In the inpatient setting, 62.0% of all pain measurements were complete and correct ([CI = 54.1, 69.5], 94.0% complete and thereof 66.0% correct, respectively) compared to 39.0% ([CI = 29.8, 48.7], 77.0% complete, and thereof 50.6% correct, respectively) of saliva samples. Overall, 52.8% [CI=46.6, 58.9] of all data collected by caregivers in the inpatient setting was complete and correct.

Table 3. Completeness and correctness of caregiver collected data and samples

			Complete and correct				
			No			Yes	
Location	**Item**	**Total n**	**n**	**%**		**n**	**%**
At home	Pain scale	153	43	28.1		110	71.9
	Saliva samples	102	47	46.1		55	53.9
	All	255	90	35.3		165	64.7
Inpatient	Pain scale	150	57	38.0		93	62.0
	Saliva samples	100	61	61.0		39	39.0
	All	250	118	47.2		132	52.8
All	Pain scale	303	100	33.0		203	67.0
	Saliva samples	202	108	53.5		94	46.5
	All	505	208	41.2		297	58.8

Reasons for incompleteness and incorrectness and exploratory sensitivity analyses

Incompleteness of data was mostly due to technical issues which two caregivers experienced (predominantly affecting saliva samples) or the discontinuation of data collection by one caregiver at home. Exploratory analyses of the subgroup of participants who did not experience technical issues with an early version of the

mobile application (n=15) showed that completeness and correctness of pain assessments remained the same (71.9%), but that the percentage of correct and complete saliva samples increased from 53.9% to 61.1%. The reason for this was a programming issue, which affected the entry of saliva sample IDs in the application.

The major reason for incorrectness of data was incorrect timing (i.e. data was not collected within predefined timeframes of 1 hour (+/-15 minutes), and four hours (+2 hours) after first medication, Table 1 in Additional file 2). Exploratory sensitivity analyses assuming that all data had to be collected within one calendar day instead of the narrow timeframes of 1 and 4 hours showed that in this case 92.2 % of pain scale data and 74.5% of saliva samples would have been complete and correct at home, and 94.0% and 73.0% in the inpatient setting, respectively (Table 2 in Additional file 2). Further, a positive trend for complete and correct data and sample collection was observable from day one in the inpatient setting to day three, i.e. the day of hospital release. On the first day at home, the proportion of complete and correct data increased once more, before then declined slightly (Additional file Figure 1). No clear trend in the quality of data collection could be identified regarding the individual patient and caregiver (Additional file Figure 2), or when stratifying the analysis by number of siblings, caregiver occupation, or caregiver native language.

Legibility of concomitant medication images

Of the 17 patients who completed the study, 10 (58.8%) took 24 different images of medications using the mobile application's imaging function. All 24 images were sharp and legible. Medication names were identifiable on 18 (75%) and dosage information (e.g. on drug containers, blisters, etc.) on 13 (54.2%) of all 24 images, respectively.

Feasibility and practicability for caregivers

Out of 17 caregivers, 15 provided answers in the feasibility questionnaire. Nine of 15 (60.0%) caregivers thought studies at home are a good idea and 53.3% (8/15) would probably take part again (Table 3 in the Additional file 2). Over 66% (10/15) of caregivers spent less than 15min on study procedures. In 73.3% (11/15), the mother was the primary caregiver collecting data and samples for the study at home. 60% (9/15) of caregivers said that study goals were explained "very well" to them in the beginning, compared to 53% (8/15) who said study procedures were explained "well". Usability of the study app was rated between "ok" (8/15, 53.3%) and "great" (6/15, 40.0%). 66.7% (10/15) rated the study procedures "easy" in the inpatient setting

compared to 40.0% (6/15) in the at home setting. Asking caregivers about potential difficulties, 86.6% (13/15) answered that the study flow, i.e. the timing of data and sample collection, was sometimes difficult to follow (Table 4 in Additional file 2).

Study cost and cost comparison

Total study cost included fixed costs ($44'577) such as application development and support ($11'955), study-specific materials (two iPods, saliva sampling tubes, envelopes, cafeteria vouchers for caregivers, $2'507), laboratory sample analysis ($2'224), study-material transport costs from caregiver's home at the end of the study ($474), database setup, management and statistical analysis ($21'360), study monitoring ($6'057) and variable salaried on site staff cost (part time; one physician, three study nurses) over the eight months the study was active ($ 19'157), summing to a total of $63'734 and $3'749 per patient who completed the study.

Our sensitivity analyses for a traditional, hospital-based approach for the same study with the same duration, but six full study days in the inpatient setting and data collection by study nurses suggested total fixed costs of $35'593 compared with $44'577 for the pilot study, and variable on-site staff costs of $20'202 compared with $19'157 for the mobile study summing to a total of $55'795 and $3'281 per patient. The difference was driven by the high initial cost for app development and support, but lower study nursing, physician, and data entry time compared to a traditional trial. Increasing the sample size from 17 to hypothetical 100 evaluable patients would have resulted in cost per patient of $1'077 for the mobile trial and $1'307 for the hospital-based approach.

Discussion

Results from our pilot study indicate that mobile data and sample collection for clinical studies with children and their caregivers are feasible, yet subject to certain caveats. We were able to engage and enrol patients and to conduct the study with retention rates comparable to those of studies done in traditional settings. Furthermore, the participating caregivers reported high levels of satisfaction and willingness to participate in similar trials in the future. However, the overall proportion of complete and correct data collected in the current framework would not be sufficient to obtain valid study results for a stand-alone PK study. However, sparse PK data collected in such at-home study may be combined with data from more

conventional PK studies to enhance PK/PD analyses including pharmacometric modelling. While 92.2% of pain scale data and 77.5% of samples were complete, only 78.0% and 69.6% thereof were correct, respectively. We could therefore not prove our hypothesis of 90% complete and correct data and sample collection in the at home setting. Reasons for this included the narrow timeframe in which data and samples had to be collected by caregivers, the handling of saliva samples, and technical issues with the application.

Expanding the narrow timeframes to one full calendar day, however, would have resulted in over 92% of pain measurements to be complete and correct in the at home setting. We therefore believe that other study types such as phase III or IV studies or observational research with less time-critical data to be collected may be viable options to make use of mobile data collection. Examples may include postoperative observations, the assessment of quality of life outcomes, medication management in chronic conditions or continuous physiological measures using sensor devices.

Exploratory analysis of factors such as number of siblings, caregiver occupation, or caregiver native language did not reveal any clear trend in supporting complete and correct data collection among caregivers. As expected, this pilot study did not yet prove cost-effective due to the development cost of the application. However, future studies including larger samples sizes and building on an improved framework of the existing application will be a cost-efficient option.

Although our participants seemed broadly similar to those in comparable traditional trials, the requirement for ease in mobile phone handling (in our pilot restricted to iPhone and iPods), understanding of the German language, and the active choice by caregivers to conduct the study at home probably may have resulted in selection bias. While this study planned to leverage study participants' own Internet-enabled mobile devices for remote data collection ("Bring Your Own Device"), we also provided study mobile devices in order to avoid additional caregiver selection bias. External validity is also a common problem for traditional trials, and adequate description of the setting and the sample characteristics is needed. In future mobile trials, we therefore aim to extend the pilot setting to different mobile technologies (i.e. Android) and multiple languages.

Furthermore, the mobile application was designed with the highest user flexibility and –usability in mind in order to minimize user fatigue or dropout. This flexibility in data entry resulted in data points which were often ambiguous (e.g. inconsistencies in automatic time stamp versus time point indicated by caregiver, or typing errors for sample codes). We decided to strictly analyse the data as transmitted by the caregivers although some of these data point ambiguities could have been resolved by the investigators by applying logical cross-checks.

While this pilot was efficient with respect to direct caregiver-reported data entry and on-site staffing levels, data management and analysis of the app-collected data structure was more resource intense than expected. Particular hurdles were (i) the translation of structured data recorded by the mobile application (json-format) into a tabular format suitable for statistical analysis and (ii) to incorporate the flexibility given to the users for data entry. For future studies, translation of structured data can be easily optimized by establishing standardized procedures, whereas specific care must be taken to reduce flexibility in data entry to avoid a high degree of complexity in data analysis.

According to the US Food and Drug Administration, electronic capture of clinical trial source data is nowadays preferred over paper-based data collection [21, 25]. However, data quality has been reported to be problematic [8, 9], and combinations of mobile technologies with appropriate interactive guidance by study staff were suggested to be more successful [3, 4]. To our knowledge, this is the first study explicitly evaluating the quality of data resulting from mobile data capture in a setting imitating pediatric PK/PD modelling. We combined an initial caregiver training session in the inpatient setting including the possibility to interact with study staff with an independent at-home phase. Caregivers were satisfied with study staff support in the inpatient setting, but reported more difficulties following the study procedures at home (Table 3 in Additional file 2).

As described by Murray [9] and Coons [25], there are generally two concerns about data quality in clinical studies: Validity – to what extent is the information provided by participants is “true”- and amount of missing data, in terms of item nonresponse. While missing data is generally expected to be minimized with automatic reminders issued by electronic applications, patient-reported data validity may be problematic in both mobile and traditional on paper studies. In this study, we confirm that missing data is less of an issue than data validity. If we had allowed caregivers to collect data

and samples within one calendar day rather than within narrow timeframes, 92.2% of all pain measurements and 74.5% of saliva samples would have been complete and correct in the at home setting (Table 2 in the Additional file 2).

Further, self-reported patient outcomes have generally been criticized before with respect to internal validity and data quality, but are often necessary to adequately evaluate the treatment benefit provided by new interventions [25]. We used a standardised scale for pain assessment [23] by caregivers as a model for mobile data collection. Automatic time stamps for data entry aimed to improve data validity (e.g. correct timing), and allowed the study team to better judge the validity of patient-reported measurements compared to paper based data. In future versions of the application, all patient-reported assessments will therefore be validated for computer use if possible, and methods for case confirmations by study staff will be included to be able to judge the validity of data points.

A recent survey [10] across pharmaceutical companies revealed that just 37% are currently using mobile technologies in clinical trials, of which more than two-thirds (68%) are mobile apps. The primary benefit that companies see for adopting “mHealth” technologies is real-time data acquisition (36%), followed closely by increased patient compliance (30%) and improved data quality (25%). The first successfully completed fully remote Diabetes management trial, the VERKKO trial, sponsored by Sanofi reported initial results on high patient satisfaction rates, reduced study coordination activities, faster study completion, and increased patient retention rates [26]. Nevertheless, companies still have a number of concerns around the technology. Data security is the primary concern for almost a third of respondents (32%), whilst difficulty in incorporation (29%) and resistance from patients or physicians (23%) are both considerable worries. While we did not experience any of these, our study revealed that data quality may only be improved if the mobile application is supporting data collection that is a) well-structured and easy to follow for patients, b) flexible to some extent (i.e. large timeframes), but still rigid enough to assure resulting data quality (i.e. using automatic time stamps rather than patient-reported time points which are prone to error), and ideally, c) remotely monitored. Compared to traditional on paper study settings, users of mobile technologies for clinical studies should make use of their potential of real-life data monitoring and automatic time stamps that should ultimately improve overall patient-reported data validity.

Due to the particular scientific and technological issues associated with the use of mobile devices that currently inhibit their widespread use, the Clinical Trials Transformation Initiative (CTTI) is developing recommendations for managing mobile devices in clinical trials, and guiding principles to promote their inclusion [11].

Conclusion

In conclusion, mobile studies conducted at home are a feasible approach when certain circumstances are met. Future electronic at-home data collection should predominantly include data that is not time critical at pre-defined, yet flexible time points. Further, the collection, storage, and shipping of biological samples have proven more difficult and should be kept at a minimum. Therefore, we would refrain from conducting remote “stand alone PK studies” that require a rigid data and samples collection schedule. However, studies with more flexible data acquisition schedules, such as clinical phase III or IV trials with sparse PK and biomarker sampling, and observational studies could profit from such an approach. Importantly, mobile applications need to be designed in a well-balanced manner, allowing for user flexibility but also assuring resulting data quality. Current efforts at the CTTI specifically target the technical and scientific issues that still remain with mobile devices in clinical research. In the future, we expect an enhanced version of our current technology to reach wider patient populations and to incorporate lessons learned with features such as remote patient monitoring using real life data capture and increased site-patient interaction.

Ethics approval and consent to participate

Ethical approval for the study “Pediatric pain management after tonsillectomy as a model to assess the feasibility of clinical studies in children in an at-home setting (TOMACHI)” was granted by the Ethics Committee Central- and North-western Switzerland (EKNZ, 2015-00022).

Acknowledgements

We thank Dr. Svetlana Beglinger for support with the design of the study and the study nurses at the University of Basel Children’s Hospital, Ms. Claudia Werner, Ms. Sabrina Trinkl, Ms. Aurora Frei, and Ms. Michelle Kress for their support with study conduct.

References

[1] Baquet CR, Commiskey P, Daniel Mullins C, Mishra SI. Recruitment and participation in clinical trials: Socio-demographic, rural/urban, and health care access predictors. Cancer Detection and Prevention. 2006;30:24-33.
[2] Heinrichs N, Bertram H, Kuschel A, Hahlweg K. Parent Recruitment and Retention in a Universal Prevention Program for Child Behavior and Emotional Problems: Barriers to Research and Program Participation. Prev Sci. 2005;6:275-86.
[3] Friend SH. App-enabled trial participation: Tectonic shift or tepid rumble? Sci Transl Med. 2015;7:297ed10.
[4] Silow-Carroll S, Smith B. Clinical management apps: creating partnerships between providers and patients. Issue Brief (Commonw Fund). 2013;30:1-10.
[5] Dorsey ER, Venuto C, Venkataraman V, Harris DA, Kieburtz K. Novel Methods and Technologies for 21st-Century Clinical Trials: A Review. JAMA Neurol. 2015;72:582-8.
[6] McAlindon T, Formica M, Kabbara K, LaValley M, Lehmer M. Conducting clinical trials over the internet: feasibility study. Bmj. 2003;327:484-7.
[7] Campbell KR. An apple a day: changing medicine through technology and engagement. Future Cardiol. 2015;11:259-60.
[8] Anguera JA, Jordan JT, Castaneda D, Gazzaley A, Arean PA. Conducting a fully mobile and randomised clinical trial for depression: access, engagement and expense. BMJ Innov. 2016;2:14-21.
[9] Murray E, Khadjesari Z, White IR, Kalaitzaki E, Godfrey C, McCambridge J, et al. Methodological challenges in online trials. J Med Internet Res. 2009;11:e9.
[10] Clinical Trials Innovation: MHealth in Clinical Trials Report: Just 37% of companies utilize technologies. 2017.
[11] Clinical Trials Transformation Initiative: Program: Mobile Clinical Trials (MCT), Mobile Devices.
[12] Shirkey HC. Therapeutic orphans--everybody's business. Ann Pharmacother. 2006;40:1174.
[13] Turner S, Nunn AJ, Fielding K, Choonara I. Adverse drug reactions to unlicensed and off-label drugs on paediatric wards: a prospective study. Acta Paediatr. 1999;88:965-8.
[14] Avenel S, Bomkratz A, Dassieu G, Janaud JC, Danan C. [The incidence of prescriptions without marketing product license in a neonatal intensive care unit]. Arch Pediatr. 2000;7:143-7.
[15] Choonara I. Unlicensed and off-label drug use in children: implications for safety. Expert Opin Drug Saf. 2004;3:81-3.
[16] Roberts R, Rodriguez W, Murphy D, Crescenzi T. Pediatric drug labeling: improving the safety and efficacy of pediatric therapies. Jama. 2003;290:905-11.
[17] Rodriguez W, Selen A, Avant D, Chaurasia C, Crescenzi T, Gieser G, et al. Improving pediatric dosing through pediatric initiatives: what we have learned. Pediatrics. 2008;121:530-9.
[18] Baiardi P, Giaquinto C, Girotto S, Manfredi C, Ceci A. Innovative study design for paediatric clinical trials. Eur J Clin Pharmacol. 2011;67:109-15.
[19] Laughon MM, Benjamin DK, Jr., Capparelli EV, Kearns GL, Berezny K, Paul IM, et al. Innovative clinical trial design for pediatric therapeutics. Expert Rev Clin Pharmacol. 2011;4:643-52.
[20] Bagust A, Haycox A, Sartain SA, Maxwell MJ, Todd P. Economic evaluation of an acute paediatric hospital at home clinical trial. Arch Dis Child. 2002;87:489-92.
[21] US Food and Drug Administration. Guidance for Industry: patient reported outcome measures: use in medical product development to support labeling claims. 2009.

[22] Baugh RF, Archer SM, Mitchell RB, Rosenfeld RM, Amin R, Burns JJ, et al. Clinical practice guideline: tonsillectomy in children. Otolaryngol Head Neck Surg. 2011;144:S1-30.
[23] von Baeyer CL. Children's self-reports of pain intensity: scale selection, limitations and interpretation. Pain Res Manag. 2006;11:157-62.
[24] Schmerzbehandlung am UKBB (Guidelines for pain management at UKBB), Version 01.2015. In: UKBB, editor.
[25] Coons SJ, Eremenco S, Lundy JJ, O'Donohoe P, O'Gorman H, Malizia W. Capturing Patient-Reported Outcome (PRO) Data Electronically: The Past, Present, and Promise of ePRO Measurement in Clinical Trials. Patient. 2015;8:301-9.
[26] eClinicalHealth Announces Successful Results for an Entirely Remote Online Clinical Trial. 2016.

Additional files

1 - Development of mobile application

A custom mobile application (“app”) was developed to allow participants to collect study information in their daily routine as easily as possible. The planned study workflow was extensively discussed with the study team and respective assumptions about the ideal app flow and different input methods were explored using multiple paper-based interactive prototypes [1]. Subsequently, a high fidelity app design mockup [2] was created in agreement with the communication department of the involved hospital. User interaction was tested again internally using clickable design prototypes [3].

Due to the required functionality, the application was developed as native mobile app primarily focused on iOS devices [4-6]. The mobile app was then beta-tested in multiple cycles with various members of the study team and people not directly involved in the study. Finally, the application was made available to the study participants through download from the Apple App Store using the participants’ own iOS device or as a pre-installed app on a dedicated study iPod-Touch.

Data from the mobile application was collected using a custom developed web server (Golang Language, RethinkDB). Data transfer was designed to be actively initiated by the study participants with all data being encrypted before transfer to the central server and transferred using SSL/TLS secured connections.

References

(1) Interactive Paper-Prototypes: POP [https://marvelapp.com/pop]
(2) Mobile App Design: Sketch [https://www.sketchapp.com]
(3) Interactive Design-Prototypes: Invision [https://www.invisionapp.com]
(4) Mobile App Development: Swift 3 [https://developer.apple.com/swift]
(5) Mobile App Development: Xcode 8.2 [https://developer.apple.com/xcode]
(6) Mobile App Development: Realm Mobile Database Library for data storage [https://realm.io]

2 - Tables & Figures

Table 1. Timing of complete data and samples, by timepoint and location

			Timing correct				
			No			Yes	
Location	**Item**	**Total n**	**n**	**%**		**n**	**%**
At home	Pain scale tM	47	0	0.0		47	100.0
	Pain scale t1	47	18	38.3		29	61.7
	Saliva sample t1	41	16	39.0		25	61.0
	Pain scale t2	47	13	27.7		34	72.3
	Saliva sample t2	38	6	15.8		32	84.2
	All	220	53	24.1		167	75.9
Inpatient	Pain scale tM	47	0	0.0		47	100.0
	Pain scale t1	47	27	57.4		20	42.6
	Saliva sample t1	38	20	52.6		18	47.4
	Pain scale t2	47	21	44.7		26	55.3
	Saliva sample t2	39	17	43.6		22	56.4
	All	218	85	39.0		133	61.0
All	Pain scale tM	94	0	0.0		94	100.0
	Pain scale t1	94	45	47.9		49	52.1
	Saliva sample t1	79	36	45.6		43	54.4
	Pain scale t2	94	34	36.2		60	63.8
	Saliva sample t2	77	23	29.9		54	70.1
	All	438	138	31.5		300	68.5

Table 2. Complete and correct data and samples when timing as a correctness factor is neglected (i.e. all data collected within one calendar day are correct)

			Complete and correct				
			No			Yes	
Location	**Item**	**Total n**	**n**			**n**	**%**
At home	Pain scale	153	12	7.8		141	92.2
	Saliva sample	102	26	25.5		76	74.5
	All	255	38	14.9		217	85.1
Inpatient	Pain scale	150	9	6.0		141	94.0
	Saliva sample	100	27	27.0		73	73.0
	All	250	36	14.4		214	85.6
All	pain scale	303	21	6.9		282	93.1
	Saliva sample	202	53	26.2		149	73.8
	All	505	74	14.7		431	85.3

Table 3. Caregiver feasibility questionnaire

	n	%
What statement(s) apply in your opinion? (n=15)		
I think clinical studies at home are a good idea	9	60.0
I don't think clinical studies at home are a good idea	1	6.7
I am not sure	5	33.3
How were the aims of the study explained to you? (n=15)		
Very well	9	60.0
Well	5	33.3
Sufficiently	1	6.7
Not well	0	0.0
Not well at all	0	0.0
How were your tasks during the conduct of the study study explained to you? (n=15)		
Very well	4	26.7
Well	8	53.3
Sufficiently	3	20.0
Not well	0	0.0
Not well at all	0	0.0
How user-friendly do you rate the mobile application? (n=15)		
Great	6	40.0
Ok	8	53.3
Unusable	1	6.7
Who has mainly collected data and samples in the hospital? (n=16)		
Mother	9	56.2
Father	1	6.2
Both parents	3	18.8
Study Nurse	3	18.8
Child	0	0.0
Other	0	0.0
Who has mainly collected data and samples in at home? (n=15)		
Mother	11	73.3
Father	1	6.7
Both parents	3	20.0
Study Nurse	0	0.0
Child	0	0.0
Other	0	0.0
How feasible were the study procedures for you in the hospital? (ntot=15)		
Very easy	4	26.6
Easy	10	66.7
Difficult	1	6.7
Very difficult	0	0.0
How feasible were the study procedures for you at home? (n=15)		
Very easy	2	13.3
Easy	6	40.0
Difficult	7	46.7
Very difficult	0	0.0
How feasible was taking photos of concomitant medication for you? (n=15)		
Not used	4	26.7

Very easy	4	26.7
Easy	5	33.2
Difficult	1	6.7
Very difficult	1	6.7
How much time did you spend on study procedures per day? (n=15)		
Less than 15 minutes	10	66.7
15-30 min	4	26.6
30-45 min	1	6.7
45-60 min	0	0.0
more than 60min	0	0.0
Would you take part again in such a study? (n=15)		
Yes, for sure	4	26.7
Probably yes	8	53.3
Probably not	3	18.8
Definitely not	0	0.0

Table 4. Main difficulties experienced by caregivers during conduct of study

What were the main difficulties you experienced during the conduct of the study? (multiple choice)		
	n	**%**
It was sometimes difficult to follow the study procedures (i.e. timing of data collection)	13	86.7
There were technical issues with the mobile application	3	20.0
My child did not want to participate, i.e. provide saliva samples	3	20.0
I was not able to contact the study personnel	2	13.3
It took too much time to follow the study procedures	1	6.7
I did not experience any difficulties	1	6.7
Other	3	20.0

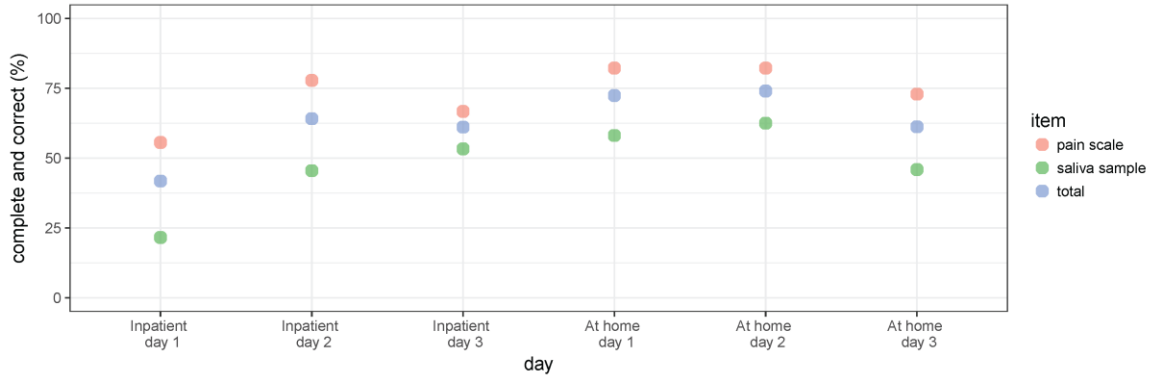


Figure 1. Proportion of complete and correct data and samples by location and day

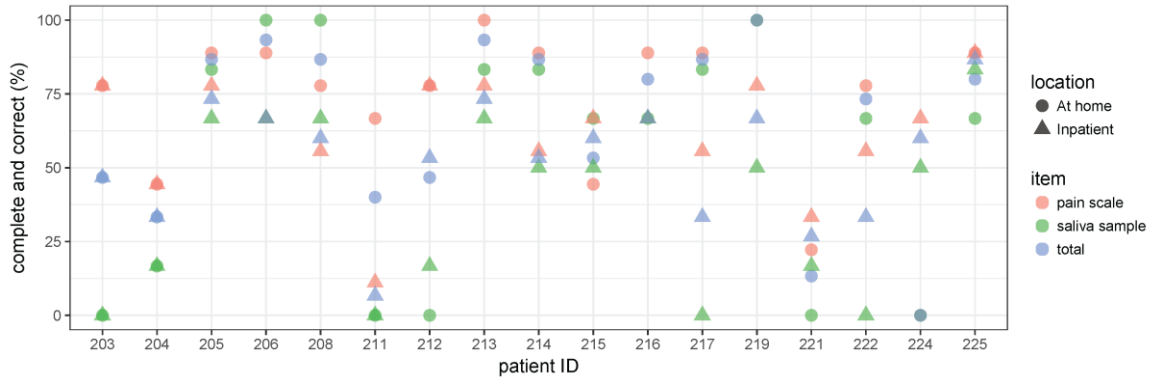


Figure 2. Proportion of complete and correct data and samples by patient and location

CHAPTER 3

FURTHER PUBLICATIONS

3.1 Published reviews (first author)

The role of Clinical Trial Units in investigator- and industry-initiated research projects.
von Niederhäusern B, Fabbro T, Pauli-Magnus C.

Swiss Med Wkly. 2015 Jul 2;145:w14161

Six multidisciplinary competence centres (Clinical Trial Units, CTUs) in Basel, Berne, Geneva, Lausanne, St. Gallen and Zurich provide professional support to clinical researchers in the planning, implementation, conduct and evaluation of clinical studies. Through their coordinated network, these units promote high-quality, nationally harmonised and internationally standardised clinical research conduct in Switzerland. We will describe why this network has been established, how it has been successful in stilling the growing need for clinical research support, which training and education opportunities it offers, and how it created national awareness for the still-existing hurdles towards clinical research excellence in Switzerland. Taking the CTU Basel as an example, we show that a considerable number (25%) of the studies submitted for regulatory approval in 2013 were supported by the CTU, decreasing the number of findings in ethics reviews by about one-third. We conclude that these achievements, together with a Swiss national funding model for clinical research, and improved national coordination, will be critical factors to successfully position Swiss clinical research at the international forefront.

3.2 Published original articles (co-author)

Premature trial discontinuation often not accurately reflected in registries: comparison of registry records with publications. Alturki R, Schandelmaier S, Olu KK, von Niederhäusern B, Agarwal A, Frei R, Bhatnagar N, Hooft L, von Elm E, Briel M.

J Clin Epidemiol. 2017 Jan;81:56-63

BACKGROUND: One quarter of randomized clinical trials (RCTs) are prematurely discontinued and frequently remain unpublished. Trial registries can document whether a trial is ongoing, suspended, discontinued, or completed and therefore represent an important source for trial status information. The accuracy of this information is unclear.

OBJECTIVE: To examine the accuracy of completion status and reasons for discontinuation documented in trial registries as compared to corresponding publications of discontinued RCTs and to investigate potential predictors for accurate trial status information in registries.

METHODS: We conducted a cross-sectional study comparing information provided in publications (reference standard) to corresponding registry entries. First, we reviewed publications of RCTs providing information on both discontinuation and registration. We identified eligible publications through systematic searches of MEDLINE and EMBASE (2010-2014) and an international cohort of 1,017 RCTs initiated between 2000 and 2003. Second, pairs of investigators independently and in duplicate extracted data from publications and corresponding registry records. Third, for each discontinued RCT, we compared publication information to registry information. We used multivariable regression to examine whether accurate labeling of trials as discontinued (vs. other status) in the registry was associated with recent initiation of RCT, industry sponsorship, multicenter design, or larger sample size.

RESULTS: We identified 173 publications of RCTs that were discontinued due to slow recruitment (55%), harm (16%), futility (11%), benefit (5%), other reasons (3%), or multiple reasons (9%). Trials were registered with clinicaltrials.gov (77%), isrctn.com (14%), or other registries (8%). Of the 173 corresponding registry records, 77 (45%) trials were labeled as discontinued and 57 (33%) provided a reason for discontinuation (of which 53, 93%, provided the same reason as in the publication). Labeling of discontinued trials as discontinued (vs. other label) in corresponding trial registry records improved over time (adjusted odds ratio 1.16 per year, confidence interval 1.04-1.30) and was possibly associated with industry sponsorship (2.01, 0.99-4.07) but unlikely with multicenter status (0.81, 0.32-2.04) or sample size (1.07, 0.89-1.29).

CONCLUSIONS: Less than half of published discontinued RCTs were accurately labelled as discontinued in corresponding registry records. One-third of registry records provided a reason for discontinuation. Current trial status information in registries should be viewed with caution.

Discontinuation and Non-Publication of Randomized Clinical Trials supported by the Main Public Funding Body in Switzerland: a Retrospective Cohort Study. Alain Amstutz, Stefan Schandelmaier, Roy Frei, Jakub Surina, Arnav Agarwal, Kelechi Kalu Olu, Reem Alturki, Belinda von Niederhäusern, Erik von Elm, and Matthias Briel

BMJ Open. 2017 Aug 1;7(7):e016216

BACKGROUND: The Swiss National Science Foundation (SNSF) is the main public funding body of randomized clinical trials (RCTs) in Switzerland. The SNSF promotes academic excellence through competitive selection of study proposals and rigorous evaluation of feasibility. Completion status and publication history of SNSF-supported RCTs have not been investigated before.

OBJECTIVES: To assess completion and publication status of all health care RCTs supported by the SNSF.

DESIGN: We established a retrospective cohort of all SNSF-supported RCTs for which recruitment and funding had ended in 2015 or earlier. For each RCT, two investigators independently searched corresponding publications in electronic databases. In addition, we approached all principal investigators to ask for additional publications and information about trial discontinuation. Teams of two investigators independently extracted details about study design, recruitment of participants, outcomes, analysis, and sample size from the original proposal and, if available, from trial registries and publications. We used multivariable regression analysis to explore potential risk factors associated with discontinuation due to slow recruitment and with non-publication, and to compare our results to data from a previous cohort of Swiss RCTs not supported by the SNSF.

RESULTS: We included 101 RCTs supported by the SNSF between 1986 and 2015. Eighty-seven (86%) principal investigators responded to our survey. Overall, 69 (68%) RCTs were completed, 26 (26%) RCTs were prematurely discontinued (all due to slow recruitment), and the completion status remained unclear for 6 (6%) RCTs. For analyzing publication status, we excluded 4 RCTs for which follow-up was still ongoing and 9 for which manuscripts were still in preparation. Of the remaining 88 RCTs, 53 (60%) were published as full articles in

peer-reviewed journals. Multivariable regression models suggested that discontinued trials were at higher risk for non-publication than completed trials (adjusted OR 7.61, 95% CI, 2.44-27.09). Compared to other Swiss RCTs, the risk of discontinuation for SNSF-supported RCTs was higher than in industry-initiated RCTs (adjusted OR 3.84, 95% CI 1.68-8.74) but not significantly different from investigator-initiated RCTs not supported by the SNSF (adjusted OR 1.05, 95% CI, 0.51-2.11). We found no evidence that the proportion of discontinued or unpublished RCTs decreased over the last 20 years.

CONCLUSIONS: One fourth of SNSF-supported RCTs were prematurely discontinued due to slow recruitment, 40% of all included RCTs and 70% of all discontinued RCTs were not published in peer-reviewed journals. There is a case to reconsider how public funding bodies such as the SNSF could improve their feasibility assessment and promote publication of RCTs irrespective of completion status.

CHAPTER 4

DISCUSSION

4.1 Improving value, reducing waste: Who's listened?

"*Perhaps all of us engaged in the enterprise we call "science" need to pause and reflect on the present state of what we do".* - After three years, I am quite confident to say that this is the centerpiece of our work described here. Our initial aim was to generate a simple cost-consequence approach to improving the quality of research. However, much more needed to be done to just even define what this equation contains. We needed to critically reflect on the current state of science, and research involving human beings in particular. This included diving into a basin of vague concepts and scarce evidence that most others had avoided swimming in so far. When we started our work, the 2009 estimate that 85% of biomedical research and more than US$ 100,000,000,000 [1] are wasted had sent ripples through the scientific community. Many causes of this waste, including asking the wrong questions, bad design, or poor publication and reporting, were called "simple problems" that could easily be fixed, for example through appropriate randomization or blinding of a clinical trial [2]. Who's listened?

As a first step, the REduce Waste and ReWARD Diligence (REWARD) campaign invited everyone involved in biomedical research to critically examine the way they work to reduce waste and maximize efficiency [2]. Institutions and researchers around the globe have signed up for the REWARD statement. One of the signees, for example, is the UK's national funding body, the National Institute for Health Research, which is committed to Adding Value in Research (AViR) by ensuring that its funded research projects answer the right questions, deliver the research efficiently and publish the results in full in an accessible and unbiased report [3]. AViR has received one of the first Cochrane-REWARD prizes that were awarded for reducing waste in research in 2017 [4]. Similarly, the Evidence Based Medicine Manifesto grew from the yearly conference "Evidence Live" at Oxford University, inviting everyone to join a movement towards better evidence by "providing a roadmap for how to achieve the listed priorities and to share the lessons from achievements already made" [5].

Since the publication of *The Lancet's* recommendations in 2014 [6-10], multiple stakeholders showed initiative to improve the system, including journals, funding agencies, and regulators [11, 12]. Still, academic institutions and the affiliated research force lagged behind. Some of the potential underlying reasons we identified throughout this dissertation was a lack of (i) a common understanding of what "quality" or "value" is constituted of, (ii) practical guidance on how to improve it in the specific setting of academic clinical research, and (iii) empirical evidence on what works, and what doesn't. I believe that we provide an answer to (i) and (ii). The impact of these findings in practice and an answer to (iii), however, will now depend on their rigorous implementation at different levels, and by different stakeholders.

4.2 The academic response

In 2014, "value" has become the preferred term over "quality". Yet, these terms are interdependent. Value is a measure of "monetary worth" and defines quality in terms of costs and prices. According to the value-based approach, a quality product is one that provides performance at an acceptable price or conformance at acceptable cost [13]. However, "*the difficulty in employing this (value-based) approach lies in its blending of two related but distinct concepts. Quality which is a measure of excellence is being equated with value, which is a measure of worth (and therefore cost). The result is a hybrid – "affordable excellence" that lacks well-defined limits and is difficult to apply in practice"* [13].

This work represents the first effort to formulate an academic response for clinical research to *The Lancet* series on increasing value, investigating both "quality" as well as "cost". Based on a systematic review of existing quality concepts, we suggest a comprehensive, consensus-based quality framework that is applicable to all study types from conceptualization of the research question to dissemination of study results. Primarily, it has been designed to be operationalized in the academic setting and fully supports the REWARD Statement [14]. At Swiss national level, the framework has triggered all stakeholders to convene in a first symposium on how to increase value of academic clinical research. In addition, we lay the foundation for future study cost assessments in academia by providing (i) a comprehensive list of items for the retrospective and prospective assessment of costs, and (ii) first evidence on main cost drivers in academic RCTs, i.e. personnel costs during the phase of trial conduct. Finally, we describe how to investigate the added value of two aspects that affect both study quality and cost, i.e. trial monitoring and data collection.

In contrast to cost, quality has always been a perceptual, conditional, and somewhat subjective attribute that may be understood differently by different people. An almost 100% agreement on a definition for quality, as realized here in our work, can therefore be seen as a major achievement. Other disciplines have a long standing and controversial tradition when it comes to defining quality, also for more simplistic things such as “products”. In 1984, David A. Garvin understood the problem, i.e. “the host of competing perspectives, each based on a different analytical framework and each employing its own terminology”, as one of coverage [13]. Scholars from philosophy, economics, marketing, and operations management had considered the subject, but each group had viewed it from different vantage points. He identified five approaches to the definition of quality: (1) the transcendent approach of philosophy, (2) the product-based approach of economics, (3) the user-based approach of economics, marketing, and operations management, and (4) the manufacturing based and (5) value-based approaches of operations management. Ideally, the approach to quality is shifted as products move from design to market (i.e. the study from conceptualization to conduct): The characteristics that connote quality must first be identified through market research (a user-based approach to quality, identifying the gap in evidence and posing a patient-relevant research question); these characteristics must then be translated into identifiable product attributes (a product-based approach to quality, design of the study and writing of protocol), and the manufacturing process must then be organized to ensure that products are made precisely to these specifications (a manufacturing-based approach to quality, conduct of the study according to protocol). Garvin however identified a common problem that is shared by each of these approaches: Each is vague and imprecise when it comes describing its basic elements [13].

With the development of our framework, we were able to comprehensively cover multiple quality dimensions and research stages that apply to all clinical studies involving patients. By formulating main quality questions for each dimension and each research stage, we offer precise guidance on how to assess quality. Moreover, it is the first framework that incorporates the patients’ voice along with other stakeholders such as funders, regulators, or industry. Although some may criticize it to be too comprehensive, bureaucratic, generalist, or theoretical, it lays the first agreed-on foundation for all stakeholders. We do not prescribe its use in practice, but we recommend covering at least all dimensions when assessing the quality of a particular study. First experiences from clinical research practice have shown that the framework is practicable and supports study staff in the development of high-quality study protocols and the internal monitoring of study conduct.

4.3 Look into the future: The value equation

At a time when the scale of investments has raised justifiable concerns, the long-run sustainability of research programs will depend on value for money. In addition to quality and the costs incurred, the value-based approach is based on the costumer's willingness to pay. While we provide guidance on the quality of research and preliminary evidence on cost drivers in this work, "willingness to pay" is a crucial subjective aspect that will need further investigation in the future. Academic decision makers in resource constrained settings will need to balance between broad improvement of quality in smaller steps and focusing on fostering excellence potential of only few projects when applying the framework. In order to be able to make these decisions, evidence on "what works, and what doesn't" is needed.

In contrast to the scarce evidence on cost drivers from industry-sponsored trials [15] identified in our systematic review, our case study on the cost of two academic RCTs should enable readers to understand how resource use and RCT costs actually correlate and how realistic cost estimates should be developed. However, our findings do not allow direct assessment of monetary investment versus quality improvement yet. The challenging future of our quality framework, therefore, will be to make it a tool to assess value for money. In the academic setting, in which resources are specifically scarce, any allocation of available funds for an improvement in quality needs to be justified. The evidence for potential quality improvement must be compelling, and worth the money. So far, the need to formally assess research value for money on funding allocation by national governments, funding agencies, and research institutions has received remarkably little recognition [16]. The framework will further need to be populated with methodological investigations on the cost-effectiveness of specific quality items in order to be able to differentiate between "nice to haves" and actual value drivers. For example, how much does patient engagement cost and how much does it eventually improve the quality of a clinical study? What is the most cost-effective approach to risk-based monitoring of a study and does it improve overall study quality? What is the most cost-effective way of collecting high quality, i.e. complete and valid, data? We tried to answer the latter two aspects by evaluating our current monitoring approach and by investigating the data quality and resulting cost-effectiveness of remote clinical studies. Surprisingly, we found that the risk-based approach ensures study data quality with much more flexibility in terms of resource use, but also leads to much more ambiguity across monitors. Similarly, while we identified a need for the conduct of patient-centered mobile studies, the resulting data quality let us question the cost-effectiveness of remote data collection. Much more evidence and many more of these studies, for example in the form of Studies-Within-A-Trial (SWATs), are therefore needed to answer questions regarding the actual value of our quality items satisfactorily.

4.4 Future directions

In the future, the aim of this work is threefold: (1) Implement, validate, and revise the current quality framework in various settings; (2) improve usability of costing tools and generate evidence on main cost drivers from different medical fields, study types, and phases; and (3) conduct research on the cost-effectiveness of individual aspects of the quality framework in order to inform decisions on actual value improvement.

In the near future, the currently ongoing initiatives to test the quality framework in real-life practice will soon generate evidence on its practicability. These findings will need to be incorporated in the adaptation and future versions of the framework. Importantly, the impact of the framework's application on study cost and quality should be rigorously monitored. First evaluations of the "Research Excellence Framework" which supports the assessment of societal and economic impact of clinical research in the UK have shown that although the proposed indicators have some validity, there are challenges in operationalizing and measuring them reliably and across different cases in a standard manner [17]. Therefore, revised versions of the framework will also need to include feedback on the feasibility and scope of reliably assessing its content. Cross-validation from different settings, e.g. different medical areas, CTUs, departments or geographic areas will support this refinement. Three pillars will be crucial to propel the framework's impact forward in the future: (1) National support from policy makers, funding bodies, regulators, methodologists, and roof organizations, (2) local buy-in from stakeholders including medical faculties, university hospitals, CTUs, and clinical investigators, and (3) broad involvement of and dissemination across patients and the public.

At Swiss national level, we plan a one-day strategy workshop to engage Swiss stakeholders, both national and local representatives, to discuss the framework's implication and implementation. The Swiss Clinical Trial Organization is currently revising its quality policies and will take the lead in implementing the framework content at local CTU level. The Office of Public Health has shown interest to consider the framework's content and definitions for the revision of the Swiss Human Research Act.

At local level, the CTUs will have to drive the implementation of the framework's content by supporting the quality of study protocols, conduct, data management, statistical analysis, or the reporting of results. This will further fuel their already existing impact on the quality of clinical research in Switzerland [18]. In our Delphi process, all CTUs were represented by their managers and quality experts who embraced the framework content and agreed on its implementation in the future. Ideally, the Departments of Clinical Research at the University

Hospitals support the CTUs and disseminate the framework content to the medical faculties of the Universities, and the affiliated research force. Researchers are particularly important to have on board, as they are at the center of research conduct. Based on the framework structure, the Department of Clinical Research in Basel is now offering a journal club that invites interested researchers and clinicians to join discussions on how to increase value of clinical research in practice.

In order to disseminate the framework content across the public and educate patients, it will be crucial to develop applications that are easily accessible, e.g. through digitalization of the framework content. The patient community around EUPATI has shown interest in teaching materials on how to identify “good research” as a research participant. A concept on how to best communicate research content to patients is currently under development at the University Hospital Basel and will serve as a driver for the development of plain language explanations of the framework. This will allow engaging patients and the public, i.e. the broadest base of citizens who have the greatest stake in research results, in the discussion on good research.

In order to generate broader evidence on the costs associated with clinical research and the main drivers, our comprehensive item list will need to be revised until it is a user-friendly tool that may be used by researchers and clinical investigators to (i) assess the cost of RCTs at their institutions retrospectively, (ii) monitor costs during the conduct of an RCT on an ongoing basis [19], and (iii) develop more comprehensive and reliable budgets before the start of the trial. In a project commissioned by the Swiss Federal Office of Public Health (FOPH), we will continue to populate our database of currently 14 retrospective cost assessments of RCTs conducted in Switzerland as described here for the case study. This will allow generating a broader evidence base on the main cost drivers stratified by medical field, study type, and phases of research. Yet, the vast diversity of studies and settings will remain a major challenge in generalizing the findings, particularly as most evidence is generated in high-income countries.

Once a solid costing methodology has been established, the framework will then need to be populated with methodological investigations on the cost-effectiveness of specific quality items. This will allow differentiating between crucial quality items, and those that can be applied depending on the setting, for example according to the risk of a study. In the long run, the use of new information technologies, such as web-based or mobile clinical trials, or routinely collected electronic health data may provide solutions to reduce costs of research while retaining or improving value [20, 21]. Alternative approaches in clinical trials, including

disease modeling, alternative trial designs, and remote assessments have already generated their results at lower costs [21-24]. Importantly, a diligent agenda should guide these transformations based on the available evidence and educate health professionals on how to design efficient and meaningful research programs. In addition to our two studies on risk-based monitoring and remote data collection, the framework will therefore also serve as guiding principle for future research-on-research at the Department of Clinical Research in Basel, and potentially other institutions.

Finally, stakeholders who are able to influence the planning, design, and conduct of academic studies, such as public funding agencies or research ethics committees, should diligently put emphasis on well-planned *a priori* feasibility assessments, well developed budgets, and other main aspects of quality that are covered in our framework. The results of the project commissioned by the FOPH on RCT costs and any future projects resulting from the work described in here should therefore be disseminated to a broad audience at policy level and be used to inform the before-mentioned stakeholders to improve practice.

4.5 Closing remarks

Value and waste have become buzzwords in the academic as well as public debate surrounding health research. Yet, evidence and guidance on the underlying concepts have been remarkably sparse. In this work, we provide a humble basis for the definition and the assessment of quality and cost of academic clinical research. We have certainly created awareness on value and waste in the academic context and engaged the major stakeholders in a discussion on how to improve the current situation. However, the impact of this work - and whether it eventually increases value in the system - now critically depends on its implementation, both nationally and on site. In whatever setting value is assessed, be it for individual studies or entire research programs, the longitudinal impact of any assessment on quality and cost will need to be rigorously monitored. Following the Lancet's recommendations, all major stakeholders have to assume responsibility for their respective parts in the value discussion. Or as Goethe would say - "Knowing is not enough; we must apply. Willing is not enough; we must do."

References

[1] Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. Obstet Gynecol. 2009;114:1341-5.
[2] REduce Waste and ReWARD Diligence (REWARD) Campaign.
[3] National Institute for Health Research: Adding value in research.
[4] Cochrane-REWARD prizes for reducing waste: 2017 Winners.
[5] Heneghan C, Mahtani KR, Goldacre B, Godlee F, Macdonald H, Jarvies D. Evidence based medicine manifesto for better healthcare. Bmj. 2017;357:j2973.
[6] Al-Shahi Salman R, Beller E, Kagan J, Hemminki E, Phillips RS, Savulescu J, et al. Increasing value and reducing waste in biomedical research regulation and management. Lancet. 2014;383:176-85.
[7] Chan AW, Song F, Vickers A, Jefferson T, Dickersin K, Gotzsche PC, et al. Increasing value and reducing waste: addressing inaccessible research. Lancet. 2014;383:257-66.
[8] Ioannidis JP, Greenland S, Hlatky MA, Khoury MJ, Macleod MR, Moher D, et al. Increasing value and reducing waste in research design, conduct, and analysis. Lancet. 2014;383:166-75.
[9] Macleod MR, Michie S, Roberts I, Dirnagl U, Chalmers I, Ioannidis JP, et al. Biomedical research: increasing value, reducing waste. Lancet. 2014;383:101-4.
[10] Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, Julious S, et al. Reducing waste from incomplete or unusable reports of biomedical research. Lancet. 2014;383:267-76.
[11] Moher D, Glasziou P, Chalmers I, Nasser M, Bossuyt PM, Korevaar DA, et al. Increasing value and reducing waste in biomedical research: who's listening? Lancet. 2016;387:1573-86.
[12] Berge E, Al-Shahi Salman R, van der Worp HB, Stapf C, Sandercock P, Sprigg N, et al. Increasing value and reducing waste in stroke research. Lancet neurol. 2017;16:399-408.
[13] Garvin DA. What Does "Product Quality" Really Mean? MIT Sloan Management Review. 1984;26.
[14] The REWARD Statement
[15] Sertkaya A, Wong HH, Jessup A, Beleche T. Key cost drivers of pharmaceutical clinical trials in the United States. Clin Trials. 2016;13:117-26.
[16] Minelli C, Baio G. Value of Information: A Tool to Improve Research Prioritization and Reduce Waste. PLoS Med. 2015;12:e1001882.
[17] Ovseiko PV, Oancea A, Buchan AM. Assessing research impact in academic clinical medicine: a study using Research Excellence Framework pilot impact indicators. BMC Health Serv Res. 2012;12:478.
[18] von Niederhausern B, Fabbro T, Pauli-Magnus C. The role of Clinical Trial Units in investigator- and industry-initiated research projects. Swiss Med Wkly. 2015;145:w14161.
[19] Eisenstein EL. DRIVE it home: making the case for prospective economic data collection in randomized clinical trials. Kidney Int. 2008;74:1507-9.
[20] Envisioning a Transformed Clinical Trials Enterprise in the United States: Establishing An Agenda for 2020: Workshop Summary. Washington DC: National Academy of Sciences.; 2012.
[21] Dorsey ER, Venuto C, Venkataraman V, Harris DA, Kieburtz K. Novel Methods and Technologies for 21st-Century Clinical Trials: A Review. JAMA Neurol. 2015;72:582-8.
[22] Wicks P, Vaughan TE, Massagli MP, Heywood J. Accelerated clinical discovery using self-reported patient data collected online and a patient-matching algorithm. Nat biotechnol. 2011;29:411-4.
[23] Jacobs BP, Bent S, Tice JA, Blackwell T, Cummings SR. An internet-based randomized, placebo-controlled trial of kava and valerian for anxiety and insomnia. Medicine (Baltimore). 2005;84:197-207.
[24] McAlindon T, Formica M, Kabbara K, LaValley M, Lehmer M. Conducting clinical trials over the internet: feasibility study. Bmj. 2003;327:484-7.

Curriculum Vitae: Belinda von Niederhäusern

Education

10.2014 - 10.2017	**University and University Hospital Basel, Basel, CH** PhD in Clinical Research**:** „Improving value of clinical research– An evidence-based approach" Supervisors: - Prof. Dr. med. Christiane Pauli-Magnus - PD Dr. med. Matthias Briel - Prof. Dr. med. Mirjam Christ-Crain - Prof. Matthias Schwenkglenks - External supervisor: Prof. Christian Burri
09.2013 - 08.2014	**The London School of Economics, London, UK** MSc International Health Policy and Health Economics
01.2012 - 08.2013	**Swiss Federal Institute of Technology (ETH), Zurich, CH** **Harvard University, Cambridge, USA** MSc Biochemistry
09.2008 - 12.2012	**Swiss Federal Institute of Technology (ETH), Zurich, CH** BSc Biochemistry

Work experience

09.2014 - 07.2015	**University Hospital Basel, Clinical Trial Unit, Basel, CH** Educational Officer Clinical Research, Postgraduate Programs
07.2014 - 09.2014	**Novartis International, ICH, Basel, CH** Intern Global Market Access
10.2013 - 07.2014	**The London School of Economics, LSE Health, London, UK** Research Assistant Health Policy & Economics Prof. Elias Mossialos
06.2013 - 08.2013	**University Hospital Basel, Clinical Trial Unit, Basel, CH** Intern Clinical Trial Management
02.2012 - 09.2012	**Swiss Federal Institute of Technology (ETH), Zurich, CH** Research Assistant Neurobiology, in vitro Prof. Martin Schwab

Awards/Selection

2014 - 2017	PPHS & SSPH+ Fellowships, fully funded PhD by CTU Basel
2016	Antelope@Novartis Mentee
2016	PPHS Top-up stipend for patient-centric research proposal
2016	Prize for second best oral presentation, Day of Clinical Research, University Hospital Basel
2015	Selected representative of Novartis BioCamp Alumni at 65th Nobel Laureate Meetings
2013	Novartis International BioCamp participant

Extracurricular activities

06.2017 - today	**Nursing Home zum Wasserturm, Basel, CH** Pro Bono Member of the Advisory Board
2015 - 2017	**University of Basel, Basel, CH** Initiator and organizer PhD Journal Club Methods in Health Sciences
2015 - 2017	**University of Basel, PhD Program in Health Sciences, CH** Doctoral students' representative
2014	**LSE Health Society Travel & Social Events Committee, UK** Travel and events coordinator

First author publications

von Niederhausern, B., T. Fabbro, and C. Pauli-Magnus, *The role of Clinical Trial Units in investigator and industry-initiated research projects.* Swiss Med Wkly, 2015. **145**: p. w14161.

von Niederhausern, B., et al., *Generating evidence on a risk-based monitoring approach in the academic setting - lessons learned.* BMC Med Res Methodol, 2017. **17**(1): p. 26.

von Niederhausern, B., et al., *Towards the development of a comprehensive framework: Qualitative systematic survey of definitions of clinical research quality.* PLoS One, 2017. **12**(7): p. e0180635.