

# Subgroup analyses in randomized clinical trials – methodological steps and pitfalls towards personalized medicine

---

***Inauguraldissertation***

*zur*

*Erlangung der Würde eines Doktors der Philosophie*

*vorgelegt der*

*Philosophisch-Naturwissenschaftlichen Fakultät*

*der Universität Basel*

*von*

*Benjamin Kasenda*

*aus Haan, Deutschland*

*Basel, 2014*

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät  
auf Antrag von:

Professor Marcel Tanner (Fakultätsverantwortlicher), Professor Heiner Bucher  
Dissertationsleiter), PD Matthias Briel (Dissertationsleiter), Professor Jos Kleijnen (Ko-  
Referent)

Basel, den 24.06.2014

(Datum der Genehmigung durch die Fakultät)

---

Prof. J. Schibler, Dekan

## Table of contents

<b>Acknowledgments</b>	<b>3</b>
<b>Plain Summary</b>	<b>4</b>
<b>Introduction</b>	<b>6</b>
<i>Subgroup analyses in randomized trials</i>	6
<i>Planning and reporting of subgroup analyses</i>	6
<i>Interactions with continuous predictors</i>	7
<i>References</i>	8
<b>Manuscripts</b>	<b>11</b>
<i>I - Learning from Failure - Rationale and Design for a Study about DISCOntinuation of Randomized Trials (DISCO study)</i>	11
Status	11
Abstract	11
Introduction	12
Study objectives and hypotheses	14
Study design and methodology	15
Data management and statistical analysis	17
Estimated sample size	18
Discussion	18
Competing interests and funding	20
Author's contribution	21
Acknowledgements	21
References	21
Tables	25
<i>II - Subgroup analyses in randomised controlled trials: cohort study on trial protocols and journal publications.</i>	29
Introduction	30
Methods	31
Results	33
Discussion	34
Conflict of interests	37
Competing interests and funding	37
Author's contribution	37
Acknowledgements	37
Ethical approval	37
Data sharing	37
References	38
Tables and Figures	40
<i>III - Investigation of Continuous Effect Modifiers in a meta-analysis on higher versus lower PEEP in patients requiring mechanical ventilation – protocol of the ICEM study</i>	46
Status	46
Abstract	46
Background	47
Objectives	47
Methods and Design	47
Discussion	52
Competing interests and funding	52
Authors' contribution	52
Acknowledgements	52
References	53

Figures	55
<i>IV - Investigation of Continuous Effect Modifiers in a meta-analysis on higher versus lower PEEP in ventilated patients with ARDS – the ICEM study</i>	56
Status	56
Abstract	56
Background	57
Patients and Methods	57
Methods for analysis	59
Results	60
Discussion	61
References	63
Tables and Figures	65
<b>Further manuscripts during PhD time</b>	<b>72</b>
<i>Published Original Articles</i>	72
Survival in Overweight Patients with Advanced Pancreatic Carcinoma: A Multicentre Cohort Study	72
Prevalence, Characteristics, and Publication of Discontinued Randomized Trials	73
<sup>18</sup> F-FDG PET Is an Independent Outcome Predictor in Primary Central Nervous System Lymphoma	74
Prognosis of patients with primary central nervous system lymphoma after high-dose chemotherapy followed by autologous stem cell transplantation.	75
Prognosis after high-dose chemotherapy followed by autologous stem-cell transplantation as first-line treatment in primary CNS lymphoma - a long-term follow-up study.	76
The prognostic value of serum methotrexate area under curve in elderly primary CNS lymphoma patients	77
<i>Original Articles accepted for publication</i>	78
Completion and publication rates of randomized controlled trials in surgery - an empirical study	78
<i>Original Articles currently under review for publication</i>	79
First-Line Treatment and Outcome of Elderly Patients with Primary Central Nervous System Lymphoma (PCNSL) – A Systematic Review and Individual Patient Data Meta-Analysis	79
<i>Published Reviews</i>	80
Lung cancer screening - an overview about chances and risks	80
Meta-analyses: what they can and cannot do.	81
<b>Discussion and Future Steps</b>	<b>82</b>
<i>Subgroup effects – trust and techniques</i>	83
<i>Future projects</i>	84
<i>Closing remarks</i>	86
<i>References</i>	87

## Acknowledgments

---

First and foremost, I would like to thank Heiner Bucher and Matthias Briel for giving me the opportunity to work at the Basel Institute for Clinical Epidemiology and Biostatistics. During this time, I had the great chance and freedom for studying, investigating, and developing. I always received support and guidance whenever needed – it was a great pleasure to work at this outstanding institute.

Thanks to Marcel Tanner for supervising my work and encouraging me to also consider the PhD time as a chance for exploring other subjects beside clinical epidemiology and taking the time to create ideas.

Thanks to Jos Kleijnen for his external review of my work and his provoking questions after my first talk about the subgroup issue at the Cochrane Colloquium in Quebec 2013 – this was the first time we met in person.

Thanks to Willi Sauerbrei and Patrick Royston for their great support with the MFPI analyses. It was an honour for me to work with such outstanding statisticians.

Thanks to all the persons involved in the DISCO-project. In particular the local site investigators Anette Blümle, Theresa Bengough, Erik von Elm, John You, and Yuki Tomonaga.

Very special thanks to Irene Oberli, the secretary of the Basel Research Ethics Committee, who always kindly supported me during the hard times of data extraction in the archives with entertaining stories and coffee.

## Plain Summary

---

Individualized or personalized medicine has become a buzzword in the academic as well as public debate surrounding health care. The word *personalized* is appealing and transports the message of a new medicine evoking hopes for patients and physicians. However, personalized medicine is not necessarily about persons - it's about subgroups and the more refined nosology of modern medicine which is based on much more profound knowledge on the pathological processes. To identify benefits and harms for these subgroups implicates several methodological issues, which I investigated in my PhD thesis.

Subgroup analyses in randomized clinical trials (RCT) can have important impact on patient care if their results are true. However, most subgroup analyses have been shown to be false with detrimental effects on patients' health. To investigate the planning of subgroup analyses in protocols of RCTs and the agreement with corresponding full journal publications, we established a cohort of RCT protocols and subsequent full journal publications. Protocols were approved between 2000 and 2003 by six research ethics committees in Switzerland, Germany, and Canada. We included 894 protocols of RCTs involving patients; 515 subsequent full journal publications were identified. About a third of protocols planned one or more subgroup analyses, but of those, only a small fraction (< 10%) provided a clear hypothesis for at least one subgroup analysis, and only a third planned an appropriate statistical test for interaction. 515 of 894 (58%) studies were published as journal article; of those, almost 50% reported at least one subgroup analysis. In 33% of all publications reporting subgroup analyses, authors stated that subgroup analyses were pre-specified but this was not supported by a third of the corresponding protocols. Furthermore, in those 86 publications in which authors claimed a subgroup effect, only 42% corresponding protocols reported a planned subgroup analysis. More than one third of statements in RCT publications about subgroup pre-specification had no documentation in the corresponding protocols. We conclude that subgroup analyses are insufficiently described in RCT protocols and investigators rarely specify the anticipated direction of subgroup effects. Credibility of claimed subgroup effects cannot be judged without access to RCT protocols.

In statistical analysis, categorizing an inherently continuous predictor (e.g. age) raises several critical methodological issues. This problem also applies to investigation of interaction between e.g. treatment assignment and a continuous predictor in RCTs – e.g. do older patients benefit from a certain therapy compared to younger patients? We applied the new multivariable fractional polynomial interaction (MFPI) approach to investigate interaction between continuous patient baseline characteristics and the allocated treatment in an individual patient data meta-analysis of 3 RCTs (N=2299) from the intensive care field. In all included RCTs, patients requiring mechanical ventilation were randomized into two treatment groups: higher *versus* lower positive end expiratory pressure (PEEP) ventilation strategy. For each study, we used MFPI to calculate a continuous treatment effect function for four baseline characteristics and 3 outcomes. These functions were plotted with a 95%

point wise confidence interval: 1. For each study separately, 2. For all studies combined (averaged function using a fixed effect model). This novel approach allows assessing whether treatment effects interact with continuous baseline patient characteristics and avoids categorisation-based subgroup analyses. These interaction analyses are exploratory in nature. However, they may help to foster future research using the MFPI approach to improve interaction analyses of continuous predictors in randomized trials and individual patient data meta-analyses.

## Introduction

---

### Subgroup analyses in randomized trials

Randomized clinical trials are the optimal design to investigate the *overall effect* of a health care intervention. The indisputable strength of this design is its ability to create patient groups that are homogenous regarding known, but also unknown prognostic factors. This allows for a relatively unbiased direct comparison between those who received the intervention and those who did not. A typical randomized clinical trial: in patients with metastatic adenocarcinoma of the lung, does therapy A compared to therapy B improve survival? Participants enrolled are selected based on defined inclusion criteria regarding patient and tumour characteristics, which are usually evaluated at the time of inclusion. Such a trial is designed to investigate the overall effect of therapy A on survival compared to therapy B – it is usually not designed to investigate whether chemotherapy A is also better in women compared to men or in patients with a worse clinical performance status at inclusion. Therefore the question arises as to whether the *overall effect* is also valid in these *subgroups*. This issue has led to many discussions and research investigating how to best apply results from randomized trials or systematic reviews to individual patient care (1-5). On the one extreme, some clinicians warn to apply overall results of large trials to individual patients or *subgroups* without considering determinants of individual effects. On the other extreme, predominantly clinical epidemiologists and statisticians warn about the danger of using subgroup results since the power is mostly not sufficient to neither show substantial benefit nor harm in respective subgroups (6, 7). In fact, subgroup analyses in RCTs or meta-analyses of RCTs are common, but their associated claims of difference of treatment effects still implicate many methodological difficulties regarding their credibility (8). Oxman and Guyatt suggested seven criteria to guide inferences about the credibility of subgroup analysis (9) and recently these recommendations were expanded to eleven criteria grouped by study design, analysis, and context (TABLE 1) (7). This approach tries to overcome the frame of absolute acceptance and rejection by placing the likelihood whether a subgroup analysis is real on a continuum from “highly plausible” to “extremely unlikely”. In other words, clinicians can judge considering each criterion: the greater the extent to which criteria are met, the more likely the subgroup effect is real. This way of appraisal mirrors the natural uncertainty about any treatment effect much better than a strict dichotomization into true or false. However, the importance of the different criteria can vary depending on the context and yet no tool has been established to assign relative weights to each criterion.

### Planning and reporting of subgroup analyses

Accumulating evidence has shown that selective reporting of results is a systemic problem afflicting all types of medical research (10). Biased reporting arises when two main decisions are made based on the direction and statistical significance of the data: (i) Should the trial results be published at all, and if so, (ii) which analyses and results should be reported in the publication. In fact, strong evidence for the selective publication of positive trials has been



available for decades (11, 12). This biased dissemination of knowledge tremendously influences further research planning, development of guidelines, and consequently the decision making in health care (10, 13). The magnitude of this so-called “file drawer” problem can only be investigated if retained study results are made available (14) or if planned analyses and endpoints are made available for researchers. These findings do not only apply for a potential *overall* effect in a clinical study, but also for subgroups regarding their pre specification, reporting, estimated effect, and analysis. The earliest stage at which a planned study is documented in detail is the study protocol submitted to a research ethics committee or a funding agency. Information from these sources is of increasing interest to methodological researchers investigating the dissemination of scientific evidence (15). As outlined above, one credibility criterion of subgroup analyses is the pre specification (7, 9, 16), which is not verifiable for readers of trial reports unless the protocol or analysis plan is available – this is usually not the case. Therefore, readers have to rely on what is reported and stated about subgroup planning. To date, there has been no empirical evidence as to which extend a statement about subgroup pre specification in a publication is trustworthy. Manuscript I entitled “*Learning from Failure - Rationale and Design for a Study about Discontinuation of Randomized Trials (DISCO study)*” describes the rationale and design of an international empirical research project in which we investigated planning, reporting, and discontinuation of RCTs – the DISCO study. Manuscript II entitled “*Subgroup analyses in randomized trials – the illusion of pre specification*” provides first empirical evidence regarding planning and reporting of subgroup analyses in RCTs based on the data acquired in the DISCO-study.

### Interactions with continuous predictors

In all branches of medical research investigators measure continuous variables e.g. age, weight, receptor expression levels on tumours, or levels of serum markers. Such continuous variables are often converted into categorical variables by grouping values into two or more categories. Various perceived advantages of dichotomizing continuous explanatory variables have been advanced, but they generally cannot be supported on statistical grounds (17). Royston and colleagues outlined that dichotomizing or categorizing an inherently continuous predictor may raise several issues for the analysis. These include the dependence of the statistical significance on the number and position of the chosen cut-point(s), possible loss of power, but also faulty interpretation if a non-linear association is incorrectly assumed to be linear (18). As an alternative, fractional polynomials (FP) have been proposed to model possible non-linearity in the relationship with the outcome of interest (19). Based on this, Royston and Sauerbrei introduced multivariable FP (MFP) modelling which is an extension that combines the selection of FP functions for several continuous variables with backward elimination of uninfluential variables. The aim of the MFP approach is to fit the data well, being simple, interpretable, and transportable (20). This concept was extended to model the interaction of continuous predictors with e.g. treatments in randomized clinical trials resulting in the multivariable fractional polynomials interaction (MFPI) approach (21, 22). Here, in a first step, MFPI estimates a fractional polynomial function representing the prognostic effect of the continuous predictor in each group, optionally adjusting for other important confounders. In a second step, the difference between the functions for the

treatment groups is calculated. A plot of the difference against the continuous predictor – treatment effect function - can be plotted which allows for qualitative assessment of the interaction with e.g. treatment allocation; such a plot is called a “treatment-effect plot.” A treatment-effect plot for a continuous covariate not interacting with treatment would be a straight line parallel to the x-axis, whereas a treatment-covariate interaction would be indicated by an increasing or a decreasing line or curve depending on the nature of association between the predictor and outcome (22). With regard to individual patient data meta-analysis, Royston and Sauerbrei recently proposed a method that allows for combining individual functions from several trials by weighted averaged estimates of a summary function (23). In cooperation with Sauerbrei and Royston I applied this novel approach to investigate interactions of continuous baseline factors based on a large individual patient data set of critically ill patients requiring mechanical ventilation. Manuscript III entitled “*Investigation of Continuous Effect Modifiers in a meta-analysis on higher versus lower PEEP in ventilated patients with ARDS – protocol of the ICEM study*” provides a detailed rationale and analysis plan. In the last manuscript IV entitled “*Continuous treatment effect modifiers in ventilated patients – ICEM study*” we report the results from our analysis.

## References

1. Akobeng AK. Understanding randomised controlled trials. Archives of disease in childhood. 2005 Aug;90(8):840-4. PubMed PMID: 16040885. Pubmed Central PMCID: 1720509.
2. Nallamothu BK, Hayward RA, Bates ER. Beyond the randomized clinical trial: the role of effectiveness studies in evaluating cardiovascular therapies. Circulation. 2008 Sep 16;118(12):1294-303. PubMed PMID: 18794402.
3. Barratt A, Wyer PC, Hatala R, McGinn T, Dans AL, Keitz S, et al. Tips for learners of evidence-based medicine: 1. Relative risk reduction, absolute risk reduction and number needed to treat. CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne. 2004 Aug 17;171(4):353-8. PubMed PMID: 15313996. Pubmed Central PMCID: 509050.
4. Dans AL, Dans LF, Guyatt GH, Richardson S. Users' guides to the medical literature: XIV. How to decide on the applicability of clinical trial results to your patient. Evidence-Based Medicine Working Group. JAMA : the journal of the American Medical Association. 1998 Feb 18;279(7):545-9. PubMed PMID: 9480367.
5. Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?". Lancet. 2005 Jan 1-7;365(9453):82-93. PubMed PMID: 15639683.
6. Rothwell PM, Mehta Z, Howard SC, Gutnikov SA, Warlow CP. Treating individuals 3: from subgroups to individuals: general principles and the example of carotid endarterectomy. Lancet. 2005 Jan 15-21;365(9455):256-65. PubMed PMID: 15652609. Epub 2005/01/18.
7. Sun X, Briel M, Walter SD, Guyatt GH. Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. Bmj. 2010;340:c117. PubMed PMID: 20354011. Epub 2010/04/01.
8. Assmann SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other (mis)uses of baseline data in clinical trials. Lancet. 2000 Mar 25;355(9209):1064-9. PubMed PMID: 10744093. Epub 2000/04/01.
9. Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. Annals of internal medicine. 1992 Jan 1;116(1):78-84. PubMed PMID: 1530753. Epub 1992/01/01.
10. Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan AW, Cronin E, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. PloS one. 2008;3(8):e3081. PubMed PMID: 18769481. Pubmed Central PMCID: 2518111. Epub 2008/09/05.

11. Dickersin K. How important is publication bias? A synthesis of available data. *AIDS education and prevention : official publication of the International Society for AIDS Education*. 1997 Feb;9(1 Suppl):15-21. PubMed PMID: 9083596. Epub 1997/02/01.
12. McGauran N, Wieseler B, Kreis J, Schuler YB, Kolsch H, Kaiser T. Reporting bias in medical research - a narrative review. *Trials*. 2010;11:37. PubMed PMID: 20388211. Pubmed Central PMCID: 2867979. Epub 2010/04/15.
13. Egger M, Smith GD. Bias in location and selection of studies. *Bmj*. 1998 Jan 3;316(7124):61-6. PubMed PMID: 9451274. Pubmed Central PMCID: 2665334. Epub 1998/02/06.
14. Blumle A, Antes G, Schumacher M, Just H, von Elm E. Clinical research projects at a German medical faculty: follow-up from ethical approval to publication and citation by others. *Journal of medical ethics*. 2008 Sep;34(9):e20. PubMed PMID: 18757621. Epub 2008/09/02.
15. Chan AW, Upshur R, Singh JA, Ghersi D, Chapuis F, Altman DG. Research protocols: waiving confidentiality for the greater good. *Bmj*. 2006 May 6;332(7549):1086-9. PubMed PMID: 16675819. Pubmed Central PMCID: 1458595. Epub 2006/05/06.
16. Yusuf S, Wittes J, Probstfield J, Tyroler HA. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. *JAMA : the journal of the American Medical Association*. 1991 Jul 3;266(1):93-8. PubMed PMID: 2046134. Epub 1991/07/03. eng.
17. MacCallum RC, Zhang S, Preacher KJ, Rucker DD. On the practice of dichotomization of quantitative variables. *Psychological methods*. 2002 Mar;7(1):19-40. PubMed PMID: 11928888.
18. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Statistics in medicine*. 2006 Jan 15;25(1):127-41. PubMed PMID: 16217841.
19. Royston P, Altman DG. Regression Using Fractional Polynomials of Continuous Covariates: Parsimonious Parametric Modelling. *Applied Statistics*. 1994;43(3):429-67.
20. Royston P, Sauerbrei W. Building multivariable regression models with continuous covariates in clinical epidemiology--with an emphasis on fractional polynomials. *Methods of information in medicine*. 2005;44(4):561-71.
21. Royston P, Sauerbrei W. A new approach to modelling interactions between treatment and continuous covariates in clinical trials by using fractional polynomials. *Statistics in medicine*. 2004 Aug 30;23(16):2509-25. PubMed PMID: 15287081. Epub 2004/08/03.
22. Royston P, Sauerbrei W. Interactions between treatment and continuous covariates: a step toward individualizing therapy. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2008 Mar 20;26(9):1397-9. PubMed PMID: 18349388.
23. Sauerbrei W, Royston P. A new strategy for meta-analysis of continuous covariates in observational studies. *Statistics in medicine*. 2011 Dec 10;30(28):3341-60. PubMed PMID: 21953493. Epub 2011/09/29.

**Table 1:** Criteria to assess the credibility of subgroup analyses as proposed by Sun et al.

<b>Design</b>
Is the subgroup variable a characteristic measured at baseline or after randomisation?
Is the effect suggested by comparisons within rather than between studies?
Was the hypothesis specified a priori?
Was the direction of the subgroup effect specified a priori
Was the subgroup effect one of a small number of hypothesised effects tested?
<b>Analysis</b>
Does the interaction test suggest a low likelihood that chance explains the apparent subgroup effect?
Is the significant subgroup effect independent?
<b>Context</b>
Is the size of the subgroup effect large?
Is the interaction consistent across studies?
Is the interaction consistent across closely related outcomes within the study?
Is there indirect evidence that supports the hypothesised interaction (biological rationale)?

## Manuscripts

---

### I - Learning from Failure - Rationale and Design for a Study about DISCOntinuation of Randomized Trials (DISCO study)

**Benjamin Kasenda\***, Erik B. von Elm\*, John You, Anette Blümle, Yuki Tomonaga, Ramon Saccilotto, Alain Amstutz, Theresa Bengough, Jörg Meerpohl, Mihaela Stegert, Kari A. O. Tikkinen, Ignacio Neumann, Alonso Carrasco-Labra, Markus Faulhaber, Sohail Mulla, Dominik Mertz, Elie A. Akl, Dirk Bassler, Jason Busse, Ignacio Ferreira-González, Francois Lamontagne, Alain Nordmann, Rachel Rosenthal, Stefan Schandelmaier, Xin Sun, Per O. Vandvik, Bradley C. Johnston, Martin A. Walter, Bernard Burnand, Matthias Schwenkglenks, Heiner C. Bucher, Gordon H. Guyatt, and Matthias Briel

\* Authors contributed equally

#### Status

**Published**, Kasenda et al. BMC Medical Research Methodology 2012, 12:131 (1)

#### Abstract

**Background:** Randomized controlled trials (RCTs) may be discontinued because of apparent harm, benefit, or futility. Other RCTs are discontinued early because of insufficient recruitment. Trial discontinuation has ethical implications, because participants consent on the premise of contributing to new medical knowledge, Research Ethics Committees (RECs) spend considerable effort reviewing study protocols, and limited resources for conducting research are wasted. Currently, little is known regarding the frequency and characteristics of discontinued RCTs.

**Objectives and Methods:** Our aims are, first, to determine the prevalence of RCT discontinuation for any reason; second, to determine whether the risk of RCT discontinuation for specific reasons differs between investigator- and industry-initiated RCTs; third, to identify risk factors for RCT discontinuation due to insufficient recruitment; fourth, to determine at what stage RCTs are discontinued; and fifth, to examine the publication history of discontinued RCTs. We are currently assembling a multicenter cohort of RCTs based on protocols approved between 2000 and 2002/3 by 6 RECs in Switzerland, Germany, and Canada. We are extracting data on RCT characteristics and planned recruitment for all included protocols. Completion and publication status is determined using information from correspondence between investigators and RECs, publications identified through literature searches, or by contacting the investigators. We will use multivariable regression models to identify risk factors for trial discontinuation due to insufficient recruitment. We aim to include over 1000 RCTs of which an anticipated 150 will have been discontinued due to insufficient recruitment.

**Discussion:** Our study will provide insights into the prevalence and characteristics of RCTs that were discontinued. Effective recruitment strategies and the anticipation of problems are key issues in the planning and evaluation of trials by investigators, Clinical Trial Units, RECs and funding agencies. Identification and modification of barriers to successful study completion at an early stage could help to reduce the risk of trial discontinuation, save limited resources, and enable RCTs to better meet their ethical requirements.

## Introduction

Randomized clinical trials (RCTs) are the optimal study design to establish the efficacy of therapeutic or preventive interventions, and are a cornerstone in drug development and comparative effectiveness research. Conducting high-quality RCTs is a challenging and resource-demanding endeavour that usually involves multiple stakeholders including clinical researchers, patients and patient interest groups, funding agencies, pharmaceutical companies, research ethics committees (RECs), and regulatory agencies.

Many unforeseen events can occur during the course of an RCT. Consequently, it is not surprising that they are often not conducted as initially planned or are prematurely discontinued.

Reasons for discontinuation of RCTs include unanticipated adverse effects (harm) (2), larger than expected benefit of an intervention (early superiority) (3), or a very low probability of detecting a designated treatment effect with continued patient recruitment or follow-up (futility) (4). RCTs may be discontinued because the sponsor withdraws funding for strategic or administrative reasons, or because new evidence from other studies may convincingly answer the primary research question or raise serious safety issues (5). Finally, RCTs are sometimes discontinued for practical reasons of insufficient recruitment of participants. To date the prevalence of trial discontinuation for any of these reasons cited above has not been determined. It also remains unknown whether the prevalence for specific reasons differs between trials initiated by investigators and those initiated by the industry.

### *Discontinued trials due to insufficient recruitment*

Difficulties in patient recruitment may necessitate amendments to the protocol. These may include prolongation of the recruitment period, broadening of inclusion criteria, addition of recruiting centres, or modifying the outcomes of interest. Some studies highlighted the high frequency of recruitment problems in RCTs (Table 1) (6-11). However, these studies only report recruitment problems of specific trials (8, 9), were based on published data (11) or the selection of trials investigated were restricted to a specific funding source (6, 7). Easterbrook et al. employed a review of study protocols (7) comparable to our approach described herein, but the data are now almost 20 years old.

Investigators have studied patients' attitudes to trial participation (12-14) and identified multiple barriers (15-17). In general, patients view clinical trials as important, ethical, and as a means of attaining superior health care for future patients. However, when asked about their own participation, responders expressed more self-concern and less altruism (12). Randomization or inclusion of a placebo arm can deter eligible patients from entering a trial (14). Other barriers to patient participation include fear of side effects, distrust of researchers, inconvenience to everyday life, complexity of protocols, fear of deterioration of the relationship with their physician, and unawareness of trial opportunities (15, 16).

In turn, attending physicians report the following barriers to an active role in trials: time constraints, lack of staff and training, worry about the impact on their relationship with patients, concern for patients, loss of professional autonomy, difficulty with the consent procedure, and lack of any reward, recognition or interest in the research question (17).

Recent research has focused on strategies of how recruitment can be improved in different settings of clinical research (18-20) and systematic reviews on the topic have identified several interventions, e.g. increasing awareness of the health problem being studied, monetary incentives, using an ‘open label’ rather than placebo design, or making trial materials culturally sensitive (21-23). Another recent systematic review emphasized the use of qualitative methods in order to identify and overcome barriers to the recruitment activity of clinicians (24). While trial discontinuation for apparent benefit has been investigated previously (25, 26), little is known about the epidemiology and features of trials discontinued for other reasons, in particular for insufficient recruitment.

#### *Ethical considerations with discontinued trials*

Trial discontinuation poses ethical problems. Firstly, study participants consent on the premise of contributing to the advancement of medical knowledge. The International Committee of Medical Journal Editors (ICMJE) argues that “patients who volunteer to participate in clinical trials deserve to know that their contribution to improving human health will be available to inform health-care decisions” (27). If trials are stopped, participants should be informed about this decision and the associated reasons. However, such information may not always be given and follow-up of already recruited participants after trial discontinuation may not always be guaranteed.

Secondly, RECs face high workloads in reviewing the protocols of planned studies. However, many RECs are under-staffed and their members serve on a voluntary basis on top of their professional duties. RECs should be enabled to identify trial projects that stand a good chance of successful completion and thereby merit the investment of a thorough review by a multidisciplinary panel. According to Article 15 of the Helsinki Declaration, RECs are also entitled to monitor the progress of approved studies (28). However, many of them may not follow up approved studies systematically despite formal requests to applicants to submit final reports or publications resulting from their research.

Thirdly, resources available for research are limited, particularly in the case of publicly funded research. Considerable waste can occur if costly RCTs need to be discontinued because assumptions about recruitment or other feasibility issues were over-optimistic (29).

Fourthly, trialists should be open about the difficulties that were encountered in failed RCTs and make their experiences available to the scientific community, in particular if the research was publicly funded. Publication of results from clinical research has been described as an “ethical imperative” (30), and in addition to data from completed studies, it has been proposed that this should also comprise information about research protocols (31). Public access to trial protocols and publication of discontinued trials is thus of high importance to help preventing replications of unsuccessful approaches and allow the inclusion of data from discontinued trials in systematic reviews. Reports of discontinued trials are available in published literature (32-35) but remain exceptions.

A comprehensive research effort using empirical methods is necessary to better understand RCT discontinuation, to meet the associated ethical challenges, and to develop guiding principles for involved stakeholders.

### Study objectives and hypotheses

We use REC-approved RCT protocols and corresponding publications to investigate the prevalence, characteristics, and publication history of RCTs that were discontinued for different reasons, and to identify risk factors for RCT discontinuation, in particular for studies discontinued due to insufficient recruitment. The specific objectives and hypotheses are:

1. To determine the risk of RCTs to be discontinued for any reason and for specific reasons including futility, adverse events, early superiority of one intervention, and insufficient recruitment (defined for primary analysis as <90% of the planned sample size achieved, and for secondary analysis as <80%).
  - Hypothesis: The prevalence of discontinued trials among approved trials ranges from 10% to 20%; insufficient recruitment of study participants is the most frequent reason for discontinuation.
2. To determine whether the risk of trial discontinuation for specific reasons will differ for investigator- versus industry-initiated trials.
  - Hypothesis: The risk for discontinuation due to insufficient recruitment is lower for industry-initiated trials.
3. To identify characteristics of study protocols associated with premature discontinuation of RCTs due to insufficient recruitment from a list of candidate variables (Table 2). These risk factors may be modifiable or non-modifiable.
  - Hypothesis: The more risk factors and the less protective factors are identified in a protocol, the higher the risk for discontinuation.
4. To determine the timing of discontinuation relative to the recruitment goals.
  - Hypotheses: a) Trials discontinued for futility are typically stopped at an advanced stage of the recruitment process (>60% of target sample size recruited); b) Trials exclusively discontinued due to insufficient recruitment are typically stopped at an earlier stage (<60% of target sample size recruited).
5. To examine the publication history of discontinued trials and to assess to what extent lessons learnt have been disseminated through formal publications, unpublished reports, databases or trial registers.
  - Hypotheses: a) Information from discontinued trials is rarely made available to others by formal publication or other forms of dissemination. b) In case of a significant result at the time of discontinuation, the results are more frequently published in a peer-reviewed journal.



## Study design and methodology

The present study addressing DISCOntinuation of RCTs (DISCO-study) is a multi-centre empirical research project that involves 4 RECs in Switzerland (Basel, Lucerne, Zurich, and Lausanne), 1 in Germany (Freiburg), and 1 in Canada (Hamilton). We have established research partnerships with each REC to access the RCT protocols approved by them between 2000 and 2003. The confidentiality of the filed study protocols is being maintained following the framework and rationale for this type of research as proposed earlier (36).

### *Eligibility criteria*

The DISCO-study is based on protocols of all approved clinical trials that allocated participants prospectively and concurrently to comparison groups by random or quasi-random methods of allocation (such as alternation, date of birth, or case record number) and compared one or several interventions with a placebo or sham intervention, another active intervention or no intervention. Studies comparing different doses or routes of administration of the same drug (early dose-finding studies), trials enrolling only healthy volunteers, or trials labeled as pilot or feasibility studies are included as pre-specified subgroups.

### *Selection process*

All study protocols approved by one of the 6 RECs between January 1<sup>st</sup> 2000 to December 31<sup>st</sup> 2002/3 will be screened for eligibility. For the purpose of the DISCO-study, we chose to sample protocols approved around 9 years ago to ensure that only a very small proportion of RCTs would be still ongoing at the time of our study (26).

### *Definition and identification of discontinued trials*

The main outcome of interest is RCT discontinuation. We define a 'discontinued RCT' as any RCT that was stopped before reaching at least 90% of the planned sample size due to any reason, including futility, adverse events (harm), early evidence of superiority of one intervention (benefit) and insufficient recruitment (a cut-off at 80% of the planned sample size will be considered in a sensitivity analysis). We use the following sources to identify discontinued trials:

- Internal REC reports on status or progress of approved studies,
- Correspondence between applicants and RECs with information about discontinuation,
- Any other specific method to identify discontinued trials used by the participating RECs,
- Any formal publication mentioning trial discontinuation,
- Directly contacting investigators about the status of the RCTs

### *Data to be extracted*

We extract data on relevant trial characteristics from protocols of eligible trials as follows:

#### Core protocol data

1. Centre and protocol information (e.g. local archive identification number, date of approval by REC)

2. Contact data of local and overall principal investigator (to enable contact with applicants through the local REC)
3. Trial properties (e.g. study design, number of centres, detailed information about interventions)
4. Trial funding (e.g. government, private for profit)
5. Any important changes/amendments to the protocol during the course of the trial (mainly extracted from correspondence between REC and applicant)
6. Main endpoints: Completion and publication status (e.g. trial stopped early for insufficient recruitment, trial published)

#### Specific protocol data

1. Clinical area (e.g. medical or surgical)
2. Setting of the trial (e.g. outpatient clinic, intensive care unit)
3. Age group of participants
4. Primary outcomes
5. Statistical analysis (e.g. planned primary analysis, intention to treat, dealing with losses to follow up)
6. Subgroups (e.g. pre-specification of subgroups)
7. Sample size, recruitment and data safety issues (e.g. planned total sample size, interim analysis, data safety monitoring board)
8. Projection of recruitment during planned enrollment time (e.g. milestones or time schedule for patient recruitment)
9. Availability of logistic/methodological support (e.g. trial support unit, structure of trial organization, paid staff at recruiting sites)
10. Strategies to support/monitor recruitment (e.g. regular newsletters, advertisement in newspapers, financial incentives)
11. Trial initiation and publication/stopping rules (e.g. industry or investigator initiated, publication constraints, sponsor rights to stop the trial)

#### *Data extraction process*

We use a web-based password-protected database (Squiekero, [www.squiekero.org](http://www.squiekero.org)) for data extraction. A manual with definitions and rules for data extraction for each variable has been compiled, updated and shared among all staff involved in data extraction at the 6 study sites. About 15 methods-trained investigators extract data from trial protocols. The course of action is illustrated in figure 1 and listed in table 3.

We conduct calibration exercises in which extracted data from several protocols will be compared and thoroughly discussed in order to ensure consistency between the investigators. This process is crucial given that some of the variables to be extracted require personal judgement. We plan to extract 30% of eligible protocols independently and in duplicate and conduct random checks for consistency in remaining protocols.

#### *Search for publications*

If no information about the publication status of a trial is given in the REC files, we conduct electronic searches in literature databases including Medline, Embase, Google Scholar, Cochrane CENTRAL register of clinical trials, CINAHL, AMED, and topic specific databases. We

also search trial registers such as ClinicalTrials.gov, ISRCTN, the WHO International Clinical Trials Registry Platform and registers of sponsors, if publicly available. We use key words from the protocol title and interventions, study acronyms, and names of the investigators as search terms. Depending on the database, we limit the searches to randomized trials in humans and take into account possible time of publication. If potential publications are found, we attempt to identify the main publication of the trial by retrieving the full text. We also check whether the main publication refers to other publications of the trial (especially rationale and design papers). From the included publications, 2 investigators extract data independently and in duplicate on the following topics: author and publication information, trial properties, study funding, clinical area, methodological quality, enrolment and follow-up, outcomes, analysis, subgroups, and sample size/recruitment.

### ***Risk factor analysis for discontinuation due to insufficient recruitment***

In a sub-study, we will compare trials that were discontinued due to insufficient recruitment with completed trials. From this subgroup, we will exclude trials that (i) used cluster randomization (because they differ from trials that randomize individuals in issues of recruitment), (ii) are still ongoing in 2012, and (iii) have unclear completion status or reasons for discontinuation other than insufficient recruitment. Trials that were discontinued due to insufficient recruitment will be considered as “cases” and all other completed trials as “controls”.

### **Data management and statistical analysis**

Data management and database cleaning will be carried out using R version 2.15.1 (The R project for statistical computing, [www.r-project.org](http://www.r-project.org)). We will read the definitive dataset into STATA (version 12.1, STATA Corporation, Austin/Texas, USA) for statistical analyses. The reasons for trial discontinuation will be analysed using descriptive statistics, including risks (cumulative incidences) of discontinuation expressed as percentage with 95% confidence intervals. In the sub-study on trial discontinuation due to insufficient recruitment, potential risk factors (hypothesis 3) will be analysed using multivariate hierarchical logistic regression models with protocol-level variables as fixed effects and the ‘participating centers’ (i.e. the RECs) as a random effect. This approach will account for variability from two sources, i.e. within and between centers. To minimize the risk of overfitting and data-driven associations, we have pre-specified risk factors and confounding variables for the statistical model and limited their number to obtain no less than 10 events (i.e. discontinued trials) per explanatory variable in the resulting multivariable logistic regression models (37).

Risk factors will include: Placebo/no treatment control *versus* active intervention, single center *versus* multicenter trial, no or inadequate *versus* adequate projection of recruitment during planned enrolment period, and absence *versus* presence of methodological/logistical support. Potential confounders will include: presence *versus* absence of industry funding/involvement, parallel *versus* cross-over/factorial trial, and the planned total number of participants.

We will calculate odds ratios with 95% confidence intervals. Statistical test results with two-sided  $P < .05$  will be regarded as significant. We expect that the proportion of missing data for the above specified variables will be low because the information to be collected from a trial protocol is either very basic or it is about the presence or absence of information in the protocol (e.g. pilot trial mentioned or not). Further, we will contact site investigators for clarifications/missing information if necessary. In our primary analysis, we will only consider protocols with complete data (complete cases analysis). In a second step, missing data will be imputed using multiple imputation techniques; based on this imputed dataset, we will conduct a sensitivity analysis (all case analysis). Furthermore, we will conduct bootstrapping for internal model validation.

### Estimated sample size

In a previous study, protocols of randomized drug trials submitted between 1989 and 1998 were analysed (38). Fifty-seven of 531 trials (11%) were discontinued for different reasons. In 22 cases (39%) the reason was insufficient recruitment of participants. In the cohort of trials established in Freiburg (Germany), 74 of 299 studies submitted in 2000 (25%) were discontinued (39). Taking into account these results and the available literature (40, 41) we estimate that about 10% to 20% of trials started are discontinued due to insufficient recruitment. Based on information by the collaborating RECs and published data, we anticipate that we will identify over 1000 eligible RCT protocols approved by the participating RECs between 2000 and 2002/3 and that about 15% of these RCTs were discontinued due to insufficient recruitment. Under the assumption of a minimal odds ratio to be detected of 2.0 and 150 of 1000 RCTs to be stopped due to insufficient recruitment, we calculated the power to detect such an association between an exposure factor (e.g. single centre status) and the binary outcome of discontinuation due to insufficient recruitment. As an example, the power to detect an association for an exposure factor is 88% if the prevalence of this factor in the “control trials” is 20% (Table 4). Therefore a sample size of 1000 protocols should be sufficient for our planned analyses.

### Discussion

The DISCO study will determine the prevalence of RCTs discontinued for a variety of reasons, differences between industry and investigator-initiated RCTs, risk factors for discontinuation due to insufficient recruitment from RCT protocols, the stage at which RCTs are discontinued, and examine the publication history of completed and discontinued RCTs. To achieve these goals a cohort of over 1000 RCTs in various medical fields will be established based on the protocols approved at participating RECs over a four-year time period. Through this publication we intend to make our study objectives and methods transparent (42).

#### *Strengths and limitations of the protocol*

In this empirical study we use robust methodology including a transparent and systematic process to identify eligible RCTs, to extract relevant characteristics from protocols, and to search for corresponding publications. The collaboration with 6 RECs in 3 different countries should enhance the generalizability of our results. Approximately 1000 RCTs will provide

sufficient statistical power for the planned analyses and likely represent one of the largest cohorts in the field of empirical trial research.

The rigor of our study depends not only on the level of detail and quality of protocols, but also on the completeness of the correspondence and amendments between the investigator and the REC. We will systematically search these files to capture any relevant information about the course of the trial, as well as on issues of recruitment or changes in design or modification of primary endpoints. In case we are not able to evaluate the completion or publication status of the trial based on the filed documents at the local REC, applicants or principal investigators will be contacted through local RECs. Experience from one of our previous projects suggests that most applicants will respond (39).

#### *Beyond discontinued trials*

The DISCO-study offers the possibility to investigate discrepancies between protocols and subsequent publications e.g. with regard to pre-specified and reported primary endpoints, statistical analyses, or sample size. As an example, judging the credibility of subgroup effects when reading trial publications is challenging and, following recent recommendations, it is crucial to pre-specify anticipated subgroup effects before the analysis (43). The DISCO-study will allow investigations about the planning and reporting of subgroup analyses in RCTs from various medical fields.

#### *Comparison with similar studies and protocols*

The STEPS study was an epidemiological survey of 114 RCTs funded by the UK Medical Research Council and Health Technology Assessment (HTA) Programme (29). Less than one-third of included trials recruited their original target number of patients within the time originally planned. Trials that reached their originally specified sample size more frequently had a dedicated trial manager, were cancer or drug trials, or offered treatments to patients exclusively available within the trial. The most commonly reported strategies to improve recruitment were newsletters and Email reminders, but the investigators could not determine whether these measures were causally linked to changes in recruitment (29).

In contrast to the STEPS study, our database will consist of RCTs that were not funded by a single agency but funded by various sponsors and sources including the industry, public, and in-house sources of university-affiliated hospitals. We will determine if the risk factors identified in the STEPS study can be reproduced within our more diverse and much larger trial cohort.

The recruitment performance of local sites within a multicentre trial is the key to successful trial completion. Recently, Dal-Ré et al. proposed the disclosure of recruitment performance of local sites within multicentre trials in publicly available trial registries (44). The rationale is that this would render the trial recruitment process more transparent and trialists more accountable, because their recruitment performance could be followed by patient organizations, sponsors, and the scientific community. The DISCO-study captures the recruitment goals of the local site and the total across all study sites, which will allow further insights into these important planning issues.

The recently finished IMPACT-study by Oude et al. (personal communication), investigated barriers and facilitators for successful patient recruitment to gynecology/obstetrics trials in the Netherlands (45). The group established a nationwide cohort of trials with recruiting physicians being interviewed about crucial determinants of recruitment at a center level. Furthermore, using a nested case-control design, they interviewed patients who refused or consented to participate in order to identify factors associated with their decision. In a second cohort study, the group investigated the association between successful recruitment and issues such as hospital organization and design of trials prospectively registered in the Netherlands Trial Register. This study, especially the latter part, has goals similar to ours. However, the methods and study population to identify risk factors are different. In IMPACT, data about potential risk factors were gathered through a questionnaire while we use data from approved protocols; and we focus exclusively on RCTs whereas IMPACT included non-randomized studies as well. The IMPACT investigators also outlined a problem regarding generalizability of potentially identified risk factors for insufficient recruitment which also applies to our protocol: on a patient level, participation or non-participation in a clinical trial might predominantly depend on characteristics of a trial and its target population; therefore overall predictors for insufficient recruitment may not be identified. We may consider this issue in sensitivity analyses e.g. through stratification by medical field. However, full data collection will demonstrate the number of events of interest; this will limit the number of variables that can be investigated in multivariable logistic regression models.

#### *Implications and significance*

The DISCO-study will provide important insights into the prevalence and features of RCTs that were discontinued for different reasons. RCTs are highly resource demanding endeavours with stakeholders including patients, clinicians, investigators, funding agencies, and industry. Effective recruitment strategies and the anticipation of problems are key issues in the planning and evaluation of trials by investigators, Clinical Trial Units, RECs and funding agencies. With the identification of potential barriers to successful study completion, the DISCO-study will help reduce the risk of premature trial discontinuation and save limited research resources. Furthermore, as outlined in the Ottawa Statement (31), RCTs imply ethical obligations to research participants. When consenting to a trial, participants accept the potential of harm that may occur to them. Their risk of harm is primarily counterbalanced by the presumed overall social good resulting from the advancement of medical knowledge. We anticipate that evidence from the DISCO-study will underpin the current efforts to enhance the transparency, standardisation and accessibility of trial information. Such improvements are crucially needed to meet the ethical obligations of RCTs and to prevent that a decline in numbers of volunteering participants will ultimately make clinical research impossible.

#### **Competing interests and funding**

This project is supported by the Swiss National Science Foundation (grant 320030\_133540/1) and the German Research Foundation (grant EL 544/1-2). JWB is funded by a new investigator award from the Canadian Institutes of Health Research and the Canadian Chiropractic Research Foundation. KAOT is supported by unrestricted grants from

the Finnish Cultural Foundation. The funding sources have no role in the design and conduct of this study and the writing of this manuscript. Depending on local politics of cooperating RECs, ethical approval for the study was either provided, or it was explicitly stated that no ethical approval was necessary.

### Author's contribution

EE, MB, and BK have designed the study and written the manuscript. They are also involved in data collection. JY, YT, AB, TB coordinate data extraction from protocols, extract data and have revised the manuscript. RS developed the web-tool for data extractions. AA, JM, MS, KT, IN, AL, MF, SM, and DM are involved in data extraction from protocols and have revised the manuscript. EA, DB, JB, IG, FL, AN, RR, SS, XS, PV, BJ, MS, and MW extract data from publications and have revised the manuscript. BB, HB, and GG supported the initiation of the study, provided logistical support, and revised the manuscript. All authors approved the final version before submission.

### Acknowledgements

We would like to thank the participating Research Ethics Committees from Germany (Freiburg), Switzerland (Basel, Lausanne, Zurich, Lucerne) and Canada (Hamilton) for their continuous support and cooperation.

### References

1. 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 medical research methodology*. 2012 Aug 28;12(1):131. PubMed PMID: 22928744. Epub 2012/08/30.
2. Barter PJ, Caulfield M, Eriksson M, Grundy SM, Kastelein JJ, Komajda M, et al. Effects of torcetrapib in patients at high risk for coronary events. *The New England journal of medicine*. 2007 Nov 22;357(21):2109-22. PubMed PMID: 17984165. Epub 2007/11/07.
3. van den Berghe G, Wouters P, Weekers F, Verwaest C, Bruyninckx F, Schetz M, et al. Intensive insulin therapy in critically ill patients. *The New England journal of medicine*. 2001 Nov 8;345(19):1359-67. PubMed PMID: 11794168. Epub 2002/01/17.
4. Ware JH, Muller JE, Braunwald E. The futility index. An approach to the cost-effective termination of randomized clinical trials. *The American journal of medicine*. 1985 Apr;78(4):635-43. PubMed PMID: 3920906. Epub 1985/04/01.
5. Smith MR, Manola J, Kaufman DS, Oh WK, Bublely GJ, Kantoff PW. Celecoxib versus placebo for men with prostate cancer and a rising serum prostate-specific antigen after radical prostatectomy and/or radiation therapy. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2006 Jun 20;24(18):2723-8. PubMed PMID: 16782912. Epub 2006/06/20.
6. Charlson ME, Horwitz RI. Applying results of randomised trials to clinical practice: impact of losses before randomisation. *British medical journal*. 1984 Nov 10;289(6454):1281-4. PubMed PMID: 6437520. Pubmed Central PMCID: 1443545.
7. Easterbrook PJ, Matthews DR. Fate of research studies. *Journal of the Royal Society of Medicine*. 1992 Feb;85(2):71-6. PubMed PMID: 1538384. Pubmed Central PMCID: 1294885. Epub 1992/02/01.

8. Wilson S, Delaney BC, Roalfe A, Roberts L, Redman V, Wearn AM, et al. Randomised controlled trials in primary care: case study. *Bmj*. 2000 Jul 1;321(7252):24-7. PubMed PMID: 10875830. Pubmed Central PMCID: 27422. Epub 2000/06/30.
9. Foy R, Parry J, Duggan A, Delaney B, Wilson S, Lewin-Van Den Broek NT, et al. How evidence based are recruitment strategies to randomized controlled trials in primary care? Experience from seven studies. *Family practice*. 2003 Feb;20(1):83-92. PubMed PMID: 12509377. Epub 2003/01/02.
10. McDonald AM, Knight RC, Campbell MK, Entwistle VA, Grant AM, Cook JA, et al. What influences recruitment to randomised controlled trials? A review of trials funded by two UK funding agencies. *Trials*. 2006;7:9. PubMed PMID: 16603070. Pubmed Central PMCID: 1475627. Epub 2006/04/11.
11. Toerien M, Brookes ST, Metcalfe C, de Salis I, Tomlin Z, Peters TJ, et al. A review of reporting of participant recruitment and retention in RCTs in six major journals. *Trials*. 2009;10:52. PubMed PMID: 19591685. Pubmed Central PMCID: 2717957. Epub 2009/07/14.
12. Cassileth BR, Lusk EJ, Miller DS, Hurwitz S. Attitudes toward clinical trials among patients and the public. *JAMA : the journal of the American Medical Association*. 1982 Aug 27;248(8):968-70. PubMed PMID: 7097966. Epub 1982/08/27.
13. Mattson ME, Curb JD, McArdle R. Participation in a clinical trial: the patients' point of view. *Controlled clinical trials*. 1985 Jun;6(2):156-67. PubMed PMID: 4006489. Epub 1985/06/01.
14. Welton AJ, Vickers MR, Cooper JA, Meade TW, Marteau TM. Is recruitment more difficult with a placebo arm in randomised controlled trials? A quasirandomised, interview based study. *Bmj*. 1999 Apr 24;318(7191):1114-7. PubMed PMID: 10213724. Pubmed Central PMCID: 27847. Epub 1999/04/24.
15. Mills E, Wilson K, Rachlis B, Griffith L, Wu P, Guyatt G, et al. Barriers to participation in HIV drug trials: a systematic review. *The Lancet infectious diseases*. 2006 Jan;6(1):32-8. PubMed PMID: 16377532. Epub 2005/12/27.
16. Mills EJ, Seely D, Rachlis B, Griffith L, Wu P, Wilson K, et al. Barriers to participation in clinical trials of cancer: a meta-analysis and systematic review of patient-reported factors. *The lancet oncology*. 2006 Feb;7(2):141-8. PubMed PMID: 16455478. Epub 2006/02/04.
17. Ross S, Grant A, Counsell C, Gillespie W, Russell I, Prescott R. Barriers to participation in randomised controlled trials: a systematic review. *Journal of clinical epidemiology*. 1999 Dec;52(12):1143-56. PubMed PMID: 10580777. Epub 1999/12/02.
18. Embi PJ, Jain A, Clark J, Bizjack S, Hornung R, Harris CM. Effect of a clinical trial alert system on physician participation in trial recruitment. *Archives of internal medicine*. 2005 Oct 24;165(19):2272-7. PubMed PMID: 16246994. Pubmed Central PMCID: 1343501. Epub 2005/10/26.
19. Francis D, Roberts I, Elbourne DR, Shakur H, Knight RC, Garcia J, et al. Marketing and clinical trials: a case study. *Trials*. 2007;8:37. PubMed PMID: 18028537. Pubmed Central PMCID: 2212650. Epub 2007/11/22.
20. Atherton H, Banks D, Harbit R, Long L, Chadd F, Hay P, et al. Recruitment of young women to a trial of chlamydia screening - as easy as it sounds? *Trials*. 2007;8:41. PubMed PMID: 18053199. Pubmed Central PMCID: 2212649. Epub 2007/12/07.
21. Caldwell PH, Hamilton S, Tan A, Craig JC. Strategies for increasing recruitment to randomised controlled trials: systematic review. *PLoS medicine*. 2010;7(11):e1000368. PubMed PMID: 21085696. Pubmed Central PMCID: 2976724. Epub 2010/11/19.
22. Watson JM, Torgerson DJ. Increasing recruitment to randomised trials: a review of randomised controlled trials. *BMC medical research methodology*. 2006;6:34. PubMed PMID: 16854229. Pubmed Central PMCID: 1559709. Epub 2006/07/21.
23. Treweek S, Pitkethly M, Cook J, Kjeldstrom M, Taskila T, Johansen M, et al. Strategies to improve recruitment to randomised controlled trials. *Cochrane database of systematic reviews*. 2010 (4):MR000013. PubMed PMID: 20393971. Epub 2010/04/16.



24. Fletcher B, Gheorghe A, Moore D, Wilson S, Damery S. Improving the recruitment activity of clinicians in randomised controlled trials: a systematic review. *BMJ open*. 2012;2(1):e000496. PubMed PMID: 22228729. Pubmed Central PMCID: 3253423. Epub 2012/01/10.
25. 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 : the journal of the American Medical Association*. 2010 Mar 24;303(12):1180-7. PubMed PMID: 20332404. Epub 2010/03/25.
26. Montori VM, Devereaux PJ, Adhikari NK, Burns KE, Eggert CH, Briel M, et al. Randomized trials stopped early for benefit: a systematic review. *JAMA : the journal of the American Medical Association*. 2005 Nov 2;294(17):2203-9. PubMed PMID: 16264162. Epub 2005/11/03.
27. De Angelis C, Drazen JM, Frizelle FA, Haug C, Hoey J, Horton R, et al. Clinical trial registration: a statement from the International Committee of Medical Journal Editors. *Annals of internal medicine*. 2004 Sep 21;141(6):477-8. PubMed PMID: 15355883. Epub 2004/09/10.
28. WMA. Declaration of Helsinki. Ethical Principles for Medical Research Involving Human Subjects. 2008 [20.04.2012]. Available from: <http://www.wma.net/en/30publications/10policies/b3/index.html>.
29. Campbell MK, Snowdon C, Francis D, Elbourne D, McDonald AM, Knight R, et al. Recruitment to randomised trials: strategies for trial enrollment and participation study. The STEPS study. *Health technology assessment*. 2007 Nov;11(48):iii, ix-105. PubMed PMID: 17999843. Epub 2007/11/15.
30. Pearn J. Publication: an ethical imperative. *Bmj*. 1995 May 20;310(6990):1313-5. PubMed PMID: 7773049. Pubmed Central PMCID: 2549684. Epub 1995/05/20.
31. Krleza-Jeric K, Chan AW, Dickersin K, Sim I, Grimshaw J, Gluud C. Principles for international registration of protocol information and results from human trials of health related interventions: Ottawa statement (part 1). *Bmj*. 2005 Apr 23;330(7497):956-8. PubMed PMID: 15845980. Pubmed Central PMCID: 556346. Epub 2005/04/23.
32. Blanton S, Morris DM, Prettyman MG, McCulloch K, Redmond S, Light KE, et al. Lessons learned in participant recruitment and retention: the EXCITE trial. *Physical therapy*. 2006 Nov;86(11):1520-33. PubMed PMID: 17079752. Epub 2006/11/03.
33. Lachin JM, Lan SP. Termination of a clinical trial with no treatment group difference: the Lupus Nephritis Collaborative Study. *Controlled clinical trials*. 1992 Feb;13(1):62-79. PubMed PMID: 1315665. Epub 1992/02/01.
34. Lopes G, Quesada J, Ahn E, Flores A, Ribeiro A, Rocha-Lima CM. Oxaliplatin and fixed-rate infusional gemcitabine in the second-line treatment of patients with metastatic colon cancer: final results of a Phase II trial prematurely closed as a result of poor accrual. *Clinical colorectal cancer*. 2007 Sep;6(9):641-5. PubMed PMID: 17945036. Epub 2007/10/20.
35. Guihan M, Garber SL, Bombardier CH, Durazo-Arizu R, Goldstein B, Holmes SA. Lessons learned while conducting research on prevention of pressure ulcers in veterans with spinal cord injury. *Archives of physical medicine and rehabilitation*. 2007 Jul;88(7):858-61. PubMed PMID: 17601465. Epub 2007/07/03.
36. Chan AW, Upshur R, Singh JA, Ghersi D, Chapuis F, Altman DG. Research protocols: waiving confidentiality for the greater good. *Bmj*. 2006 May 6;332(7549):1086-9. PubMed PMID: 16675819. Pubmed Central PMCID: 1458595. Epub 2006/05/06.
37. Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and Cox regression. *American journal of epidemiology*. 2007 Mar 15;165(6):710-8. PubMed PMID: 17182981. Epub 2006/12/22.
38. von Elm E, Rollin A, Blumle A, Huwiler K, Witschi M, Egger M. Publication and non-publication of clinical trials: longitudinal study of applications submitted to a research ethics committee. *Swiss medical weekly*. 2008 Apr 5;138(13-14):197-203. PubMed PMID: 18389392. Epub 2008/04/05.

39. Blumle A, Antes G, Schumacher M, Just H, von Elm E. Clinical research projects at a German medical faculty: follow-up from ethical approval to publication and citation by others. *Journal of medical ethics*. 2008 Sep;34(9):e20. PubMed PMID: 18757621. Epub 2008/09/02.
40. Haidich AB, Ioannidis JP. Effect of early patient enrollment on the time to completion and publication of randomized controlled trials. *American journal of epidemiology*. 2001 Nov 1;154(9):873-80. PubMed PMID: 11682370. Epub 2001/10/30.
41. Haidich AB, Ioannidis JP. Patterns of patient enrollment in randomized controlled trials. *Journal of clinical epidemiology*. 2001 Sep;54(9):877-83. PubMed PMID: 11520646. Epub 2001/08/25.
42. Godlee F. Publishing study protocols : making them visible will improve registration , reporting and recruitment. *BMC medical research methodology*. 2001:4-6.
43. Sun X, Briel M, Walter SD, Guyatt GH. Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. *Bmj*. 2010;340:c117. PubMed PMID: 20354011. Epub 2010/04/01.
44. Dal-Re R, Moher D, Glud C, Treweek S, Demotes-Mainard J, Carne X. Disclosure of investigators' recruitment performance in multicenter clinical trials: a further step for research transparency. *PLoS medicine*. 2011 Dec;8(12):e1001149. PubMed PMID: 22215987. Pubmed Central PMCID: 3246429. Epub 2012/01/05.
45. Oude Rengerink K, Opmeer BC, Logtenberg SL, Hooft L, Bloemenkamp KW, Haak MC, et al. IMproving PArticipation of patients in Clinical Trials--rationale and design of IMPACT. *BMC medical research methodology*. 2010;10:85. PubMed PMID: 20875119. Pubmed Central PMCID: 2955658. Epub 2010/09/30.

## Tables

**Table 1:** Examples of studies reporting about recruitment problems in randomized controlled trials (RCTs).

Authors	Year	Data Source	Findings
Charlson et al.	1984	41 RCTs ( $\geq 250$ patients) identified by an inventory of the National Institute of Health in 1979; investigator survey was principal data source	A third of RCTs recruited fewer than 75% of their planned sample size
Easterbrook et al.	1992	720 research protocols (N=137 RCTs) approved by REC (UK); investigator survey was principal data source	Main reason (28%) for terminating the study was slow recruitment of patients
Wilson et al.	2000	RCT that investigated two management strategies for dyspepsia in primary care (UK)	90 primary care physicians were contacted; 43 agreed to participate, 31 recruited at least one patient, only 23 recruited more than 5 patients.
Foy et al	2003	7 primary care trials of dyspepsia management in the UK	One study reached its recruitment target; five recruited less than 50% of target and three of those closed prematurely
McDonald et al.	2006	114 RCTs funded by the Medical Research Council and Health Technology Assessment (UK); full scientific applications and subsequent trial reports were principal data source	Less than a third of the trials achieved their original recruitment target
Toerien et al.	2009	133 publications of RCTs identified by a systematic literature review (restricted to six major journals)	Of those trials reporting sample size calculation, 21% failed to achieve planned numbers at randomisation and 48% at outcome assessment.

**Table 2:** Potential risk factors and protective factors for trial discontinuation due to slow recruitment

Modifiable factors		Non-modifiable Factors	
Risk	Protective	Risk	Protective
Burdensome data collection at recruiting sites	Support from a methods centre, clinical trials unit, or contract research organization	Placebo control	Active treatment as control
No professional staff at recruiting centres to manage the trial	Paid local staff at recruiting centres, dedicated central trial coordinator, patient involvement in trial planning and/or conduct	No external funding	Externally funded or fully Industry sponsored
No projection of recruitment rates	Projection of patient recruitment based on e.g. pilot trial applying the full protocol or other checks for eligible patient volume	Long duration of follow-up	Short duration of follow-up / High community interest in research topic (e.g. new technology or new treatment)
No consideration of recruitment strategies	Consideration of recruitment support strategies (e.g. regular visits/audits by PI; specific training held for recruiting staff; regular progress reports; posters and information leaflets etc.)	No research network, low trial experience	Experienced PI/steering committee/network of recruiting centres for RCTs
Single centre trial	Multicentre trial	Equivalence/non-inferiority design	Intervention only available through trial participation
Low motivation for recruiting sites	Financial incentives for recruiting staff and participants	Critically ill or paediatric patients as target population	Trial experience with certain vulnerable trial populations

**Table 3:** Steps for identification of discontinued trials and data extraction; REC, regional ethics committee.

Steps	Actions
1	Identification of protocols of RCTs submitted 2000 to 2002 with the help of REC staff members
2	Extraction of trial characteristics from eligible protocols and attempt to clarify completion of trials through filed correspondence between the REC and applicants
3	Electronic search for publications (e.g. MEDLINE, EMBASE, Google Scholar) of eligible trials using filed information such as key words from protocol title/intervention or names of investigators
4	REC in charge will contact the applicants using a standardized questionnaire to ask about reasons of discontinuation and the availability of any formal publications, unpublished reports or other information from eligible trials (only in case trial completion and publication status remain unknown after searching filed correspondence and comprehensive publication search)
5	The REC in charge may send several reminders or contact applicants by phone if necessary
6	After receiving responses from applicants the data collection process will be finalized
7	The analysis database will contain only anonymous data with trial identification numbers

**Table 4:** Power calculations for different prevalences of a single risk factor for trial discontinuation; RCT, randomized controlled trial; OR, odds ratio.

Prevalence (%) of risk factor		Study power (%) to detect OR=2.0
Completed RCTs	RCTs discontinued due to slow accrual	
10	18	65
20	33	88
30	46	95
40	57	96

## II - Subgroup analyses in randomised controlled trials: cohort study on trial protocols and journal publications.

**Benjamin Kasenda**, Stefan Schandelmaier, Xin Sun, Erik von Elm, John You, Anette Blümle, Yuki Tomonaga, Ramon Saccilotto, Alain Amstutz, Theresa Bengough, Joerg J. Meerpohl, Mihaela Stegert, Kari A. O. Tikkinen, , Ignacio Neumann, Alonso Carrasco-Labra, Markus Faulhaber, Sohail M. Mulla, Dominik Mertz, Elie A. Akl, Dirk Bassler, Jason W Busse, Ignacio Ferreira-González, Francois Lamontagne, Alain Nordmann, Viktoria Gloy, Heike Raatz, Lorenzo Moja, Rachel Rosenthal, Shanil Ebrahim, Per O. Vandvik, Bradley C. Johnston, Martin A. Walter, Bernard Burnand, Matthias Schwenkglenks, Lars G. Hemkens, Heiner C. Bucher, Gordon H. Guyatt, and Matthias Briel

BMJ. 2014 Jul 16;349:g4539

### Abstract

**Objectives.** To investigate the planning of subgroup analyses in protocols of randomized controlled trials (RCTs) and the agreement between these plans and the reporting in full journal publications.

**Design.** Cohort of RCT protocols and subsequent full journal publications

**Setting.** Six research ethics committees in Switzerland, Germany, and Canada.

**Data sources.** 894 protocols of RCTs involving patients approved by participating research ethics committees between 2000 and 2003 and 515 subsequent full journal publications.

**Results.** Of 894 RCT protocols, 252 (28.1%) planned subgroup analyses. Of those, 17 (6.8%) provided a clear hypothesis for at least one subgroup analysis, 10 (4.0%) anticipated the direction of a subgroup effect, and 87 (34.5%) planned a statistical test for interaction. Of 515 identified journal publications, 246 (47.8%) reported at least one subgroup analysis. Of those, industry-sponsored RCTs more often planned subgroup analyses in the protocols compared to investigator-sponsored trials (86/160 [54%] *versus* 28/86 [33%],  $p = 0.001$ ). In 81 (32.9%) of the 246 publications reporting subgroup analyses, authors stated that at least one subgroup analysis was pre-specified; this statement, however, could not be verified in 28 protocols (35.6%). In 86 publications authors claimed a subgroup effect, but only 36 (41.8%) corresponding protocols reported a planned subgroup analysis.

**Conclusions.** Subgroup analyses are insufficiently described in RCT protocols submitted to research ethics committees and investigators rarely provide clear hypotheses with anticipated direction of subgroup effects. More than one third of statements in RCT publications about subgroup pre-specification had no documentation in the corresponding protocols. Credibility of claimed subgroup effects cannot be judged without access to RCT protocols.

## Introduction

The primary goal of a randomized controlled trial (RCT) is to determine the benefits and harms of an intervention. However, trial populations are typically heterogeneous regarding individual patient characteristics such as age, sex, disease severity, or comorbidity. The question therefore arises as to whether intervention effects vary across these patient characteristics. RCTs commonly report exploration of such possible subgroup effects (1-5) and, if conducted appropriately, such exploration can lead to more targeted clinical recommendations, better informed clinical decision-making, and improved patient care (6, 7). More frequently, unfortunately, their results are misleading and can have detrimental consequences (8, 9).

Because subgroup analyses may be either informative or misleading, health care providers and policy makers need criteria to differentiate credible from spurious subgroup effects (8, 10).

Clinical epidemiologists have suggested criteria (8, 9, 11, 12) that allow readers to gauge the likelihood that a subgroup effect is real on a continuum from *highly plausible* to *extremely unlikely* (13). All available criteria include the pre-specification of subgroup analyses; some additionally include the anticipated direction of the subgroup effect and the use of a statistical test addressing the likelihood that apparent subgroup effects may be explained by chance (8, 9, 11-13).

Judging the credibility of a reported subgroup effect relies on the information provided in published articles, because trial protocols are usually not freely accessible. Little is known about the planning of subgroup analyses in trial protocols and the extent to which they are reported in subsequent publications, and, in particular, to which claims of pre-specification correspond to these descriptions (14, 15). Pioneer work was done by Chan et al.(16); they suggested large discrepancies between protocols and publications, but their sample was limited to 70 RCT protocols from a single centre.

We investigated subgroup planning and reporting based on RCT protocols from six international centres and corresponding publications and focused specifically on the agreement between statements about subgroup pre-specification in the publication and corresponding statements in the protocols (17).



## Methods

### *Study Design*

We used RCT protocols and corresponding publications included in a retrospective cohort study; the rationale and design has been described elsewhere (17). In short, the study examined RCT protocols approved between 2000 and 2003 by six research ethics committees in Switzerland (Basel, Lucerne, Zurich, and Lausanne), Germany (Freiburg), and Canada (Hamilton). We focused on protocols that had been approved 10 or more years ago to ensure that the number of ongoing RCTs would be limited (18).

### *Eligibility Criteria for Protocols and Subsequent Publications*

In the present study, we included RCT protocols regardless of publication status. We excluded protocols of trials that: (1) compared different doses or routes of administration of the same drug (early dose-finding studies), (2) enrolled only healthy volunteers, (3) were never started, or (4) were still ongoing as of April 2013. We included only full (peer-reviewed) journal publications from corresponding RCT protocols; research letters, letters to the editor, or conference abstracts were excluded.

### *Definitions*

We defined a subgroup as a subset of all trial participants with distinct characteristics at randomisation (e.g. age, sex, stage of disease). We defined a subgroup analysis as an analysis that explored whether intervention effects (experimental *versus* control) differed according to these characteristics. For protocols, we considered a subgroup analysis as planned if at least one of the following was reported: (1) any statement in the protocol analogous to the definition above (e.g. ‘intervention effects will be investigated according to patient baseline characteristics’); (2) a stratified analysis (e.g. ‘patients will be stratified according to sex and analysed separately’); (3) a test for interaction, i.e. interaction between intervention and patient characteristic; or (4) an investigation of intervention modifying factors. For publications, we considered a subgroup analysis as reported if the article included at least one of the following: (1) an effect estimate and an associated confidence interval or a *P*-value for one or more subgroups; (2) a difference between effect estimates of different patient subgroups; (3) investigation of an intervention modifying effect or the results from a test for interaction; or (4) an explicit statement that a subgroup analysis had been undertaken. We assessed RCT protocols for industry- or investigator-sponsorship using the following criteria: The protocol clearly named the sponsor, displayed a company or institution logo prominently, mentioned affiliations of protocol authors, included statements about data ownership or publication rights, or statements about full funding by industry or public funding agencies (18).

### *Data Extraction Process and Search for Publications*

Twelve investigators trained in clinical research methodology independently extracted data from eligible trial protocols and correspondence between the research ethics committees and the local investigators. Thirty per cent of the extractions were done in duplicate as an initial calibration process to maximize the consistency of data extraction across reviewers. If the files of the ethics committee provided no information about the publication status of a

trial, we conducted comprehensive searches of electronic databases to find any associated publications; previous publications present details of the searches and data extraction process (17, 18). When RCTs that mentioned any pre-specified subgroup analyses in their publications did not mention any subgroup analyses in corresponding protocols, we searched for additional protocol versions published in journals, any available analysis plans (from journals, REC files, or websites), and information published in trial registries (clinicaltrials.gov, WHO International Clinical Trials Registry Platform). Twenty-two investigators trained in clinical research methodology extracted data from all corresponding publications, independently and in duplicate; disagreements were resolved by consensus or by third party adjudication.

#### *Information collected about subgroup analyses*

We recorded the number of subgroup analyses planned in protocols and reported in publications. We asked the following questions guided by criteria for the credibility of subgroup analyses (19):

For protocols: Any subgroup analyses mentioned? If yes: (1) Any clear hypothesis for the planned subgroup analyses mentioned? (2) Any anticipated direction of a subgroup effect mentioned? (3) Any test for interaction mentioned? (4) How many subgroup analyses were planned?

For publications: Does the publication report any subgroup analysis? If yes: (1) Does the publication report that subgroup analyses were pre-specified? (2) Does the publication report that subgroup analyses were done *post hoc*? (3) Does the publication provide a rationale for any subgroup analysis? (4) Does the publication report an anticipated direction of any subgroup effect? (5) Does the publication report any separate power calculation for subgroup analyses? (6) Does the publication report any test for interaction? (7) How many subgroup analyses are reported? (8) Does the publication report any claim about a subgroup effect? We considered a subgroup effect as claimed if the investigators explicitly state in the abstract or discussion/conclusion that the effect of an intervention was different between subgroups or a clear benefit/harm was seen in one or more subgroups.

#### *Statistical Considerations*

For binary data we summarized results as frequencies and proportions and for continuous data as medians and interquartile ranges. We considered three analysis sets: (1) a dataset based on all protocols (protocol set), (2) a dataset based on corresponding publications (publication set), and (3) a dataset of publications and matched corresponding protocols (publication-protocol set). We pre-specified stratification of our descriptive analyses by sponsorship and hypothesized based on results reported by Sun et al. that industry-sponsored RCTs more often planned subgroup analyses (1). This difference between proportions was statistically examined using the Chi-squared test. We used the statistical programmes R version 2.15.3 ([www.r-project.org](http://www.r-project.org)) and STATA version 13.0 (Stata Corp, College Station, TX, USA) for our analyses.

## Results

### *Planning of subgroup analyses – the protocol set*

Of 894 eligible RCT protocols involving patients (FIGURE 1), 252 (28.2%) planned at least one subgroup analysis. Those RCTs planning subgroup analysis had on average a larger sample size, were more frequently multicentre trials, industry-sponsored, and from the cardiovascular field (TABLE 1). Of the 252 RCT protocols planning at least one subgroup analysis, 17 (6.7%) provided a hypothesis and 10 (4.0%) provided an anticipated direction of a potential subgroup effect (TABLE 2). Interaction tests were planned in 87 (34.5%) RCT protocols with no differences between sponsor types.

### *Reporting of subgroup analyses – the publication set*

For 515 RCT protocols we identified corresponding full journal publications (publication set, FIGURE 1). Of those, 246 (47.8%) publications reported subgroup analyses. These RCTs were, on average, larger and more often published in high impact journals than published RCTs without subgroup reporting (ONLINE TABLE 1). TABLE 3 summarizes the reporting of subgroup credibility criteria and characteristics of subgroup analyses in these full journal publications. Similar to the protocol set, subgroup hypotheses or anticipated directions of subgroup effects were rarely provided. Of 86 publications claiming a subgroup effect, 39 (45.3%) reported the use of an interaction test, 9 (10.5%) provided a subgroup hypothesis, and 5 (5.8%) provided an anticipated direction of effect.

### *Agreement between subgroup reporting in publications and corresponding protocols - the publication-protocol set*

Of 246 publications that reported subgroup analyses, 114 (46.3%) corresponding protocols planned at least one subgroup analysis. In those 114 RCTs, the reported number of subgroup analyses matched the planned number in the protocol in 11 (9.6%) instances. Agreements of subgroup credibility criteria for those 246 trials reporting at least one subgroup analysis are summarized in TABLE 4. In 81 of 246 (32.9%) publications reporting subgroups, authors stated for at least one of their reported subgroup analyses that it was pre-specified, but 28 (34.6%) corresponding protocols had not mentioned any planned subgroup analysis. For 12 of these 28 RCTs, the authors mentioned a separate analysis plan in the publication or the protocol without mentioning subgroup analyses. However, these analysis plans were not made available to readers. We found registered information for 9 (32.1%) of the 28 RCTs but without any evidence of planned subgroup analyses. Of the 86 publications claiming a subgroup effect, 36 (41.8%) corresponding protocols reported a planned subgroup analysis.

## Discussion

### *Principal findings*

Our study provides empirical evidence documenting the planning and reporting of subgroup analyses in a sample of 894 patient RCTs approved by six research ethics committees in three countries. About half of published RCTs reported the conduct of subgroup analyses, of which only 46% mentioned any planned subgroup analyses in the corresponding protocols. Industry-sponsored RCTs planned subgroup analyses more often than investigator-sponsored trials, but still only half of industry-sponsored trials reporting subgroup results explicitly stated such planned analyses in the protocol. In trials with subgroup analyses mentioned in both the protocol and the publication, the number of subgroup analyses reported in publications matched the number in protocols in only 10%. Investigators rarely provided a rationale for or indicated the anticipated direction of potential subgroup effects in either protocols or RCT reports. Of 81 journal publications stating that at least one subgroup analysis was pre-planned, a third failed to mention any subgroup analysis in the corresponding protocol.

### *Strengths and limitations*

The data for the present study were collected as part of a large international cohort involving six research ethics committees that allowed full access to trial protocols and filed correspondence. As outlined previously (20), unrestricted access is absolutely necessary (but not always granted) to maintain scientific rigor: asking trialists and sponsors for permission to access their protocols would very likely introduce bias, because those with substandard reporting practices may be less likely to allow additional scrutiny. As a further strength we involved only trained methodologists in data abstraction. Finally, our sample included RCTs from various fields of clinical medicine thus enhancing generalizability of our results.

Our study has limitations. First, we did not have access to statistical analysis plans that may have pre-specified subgroup analyses not mentioned in the protocol. However, we exhaustively checked all available evidence (published protocols, trial websites, REC files, trial registries) for pre-specification of subgroup analyses. Second, we did not systematically extract information from protocols about separate power calculations for subgroup analysis. However, since only 4% of protocols that planned subgroup analysis provided an anticipated direction of a subgroup effect, we estimate that appropriate power calculations (additionally including an estimate for the magnitude of the subgroup effect) was likely less frequent than 4%. Only 2.4% of publications that mentioned a subgroup analysis reported a corresponding power calculation. Third, we used a convenience sample of six research ethics committees, which were – to our knowledge – not in any way particular. Still, we cannot say whether they are representative for other research ethics committees in their own or other countries. Fourth, due to limited resources we used single data extraction for almost 70% of protocols, thereby potentially increasing extraction errors. However, we used pre-piloted extraction forms with detailed written instructions, conducted formal calibration exercises with all data extractors, and checked extractions from a random sample of protocols at several points during the process. Agreement was good with no more than 2 discrepancies in 30 extracted answers. All data extractions from identified publications were performed in duplicate (21). Protocols and corresponding publications were not extracted by the same

person. Fifth, instead of a formal protocol we previously published only a protocol of the overall project mentioning this study without giving details (17). However, we kept the analyses of this study descriptive except for one prespecified subgroup hypothesis. We included our data extraction forms reflecting all collected variables. Sixth, included RCT protocols were approved 10-13 years ago; the planning of subgroup analyses in protocols may have improved since that time.

#### *Comparison with other studies*

In an earlier systematic review of 469 RCTs (19) we found that 44% of full text RCT publications reported subgroup analyses, which is consistent with our finding of 48% in the present study. In the prior study, we found that most claimed subgroup effects in RCTs had low credibility and pre-specification was seldom reported. The present study not only confirms this finding, but reveals that when articles claim pre-specification of subgroups, about a third of the corresponding protocols fail to mention the pre-specification.

Many previous empirical studies complained that justification of subgroup analysis and the statistical methods used were very rarely reported (2-5, 14, 16, 22, 23). However, only a few smaller studies (samples of 37 to 70 RCTs) compared subgroup analyses outlined in grant applications (23) or RCT protocols (14, 16) with reported subgroup analyses in publications. All reported a high frequency of discrepancies: Boonacker et al. noted that only 11 of 47 (23%) grant proposals for RCTs were in agreement with publications (23); Chan et al. found that 25 of 70 (36%) RCTs reported subgroup analyses in the protocol or in the publication and that there were discrepancies between the two documents in all 25 RCTs (16); Al-Marzouki et al. documented that only 8 of 19 (42%) RCT protocols not mentioning subgroup analyses and 7 of 18 (39%) RCT protocols planning subgroup analyses were in-line with the reporting in corresponding publications (14). In our sample, less than 5% (11/246) of RCTs with subgroup analyses reported in the publication planned the same number of subgroups in protocols. Only Chan et al. examined whether reported pre-specification of subgroup analyses in publications (7/20, 35%) was backed-up by planned subgroup analyses in protocols, which was not true for 4 (57%) of 7 RCTs with reported pre-specifications (16).

#### *Implications for reporting and interpreting subgroup analyses*

Recommendations for judging credibility of subgroup analyses are intended to help readers using information provided in the publication. However, because empirical evidence from comparisons of journal publications and RCT protocols has been very limited (14, 16), our results challenge one of the criteria that all previous recommendations suggest, i.e. the *a priori* specification of the subgroup analysis. Given that in one out of three studies that claimed such pre-specification, the protocols provide no corroboration; gains in credibility from the pre-specification criterion are limited.

The following steps could help to improve the trustworthiness of reported subgroup analyses: First, if subgroup analyses are pre-specified, this should be documented in trial registries. To date, possibilities to enter planned subgroup analyses, however, are insufficiently developed in trial registries. There is a non-mandatory „Group/Cohort“ field on [clinicaltrials.gov](http://clinicaltrials.gov) that could be used for subgroup pre-specification, but the corresponding

data element description remains unclear (24). The WHO International Clinical Trials Registry Platform (25) or the Controlled Clinical Trials platform (26) do currently not provide any fields for subgroup entry. Second, clinical investigators should adhere to and research ethics committees should strictly endorse adherence to reporting guidelines of RCTs protocols (SPIRIT statement) (27, 28).

Third, journal editors should insist that trial protocols and/or statistical analysis plans are provided together with publications and made accessible to readers. In addition, journal editors should enforce adherence to guidelines for RCT reports (CONSORT statement) (29) to minimize underreporting of subgroup items. Unless a reliable source such as a comprehensive trial protocol is available, readers of publications should consider statements about subgroup pre-specifications with scepticism. Instead, when judging the credibility of a subgroup effect, they may look for similar studies potentially showing consistency of subgroup findings.

### *Conclusion*

There are large discrepancies between the planning and reporting of subgroup analyses in RCTs. Published statements about subgroup pre-specification could not be verified in about a third of cases. Our results highlight the importance of enhancing the reporting quality of RCT protocols and their accessibility.

### **Conflict of interests**

All authors have completed the ICMJE uniform disclosure form at [www.icmje.org/coi\\_disclosure.pdf](http://www.icmje.org/coi_disclosure.pdf) (available on request from the corresponding author) and declare: no financial relationships with any organisations that might have an interest in the submitted work in the previous three years, and no other relationships or activities that could appear to have influenced the submitted work.

### **Competing interests and funding**

This study was funded by the Swiss National Science Foundation (grant 320030\_133540/1) and the German Research Foundation (grant EL 544/1-2). MB, AN, VG, HR, LGH, and HCB are supported by Santésuisse and the Gottfried and Julia Bangerter-Rhyner-Foundation. XS is supported by a Young Investigators Award (2013SCU04A37) from Sichuan University, China. During study preparation, EvE was supported by the Brocher Foundation. JWB is funded by a New Investigator Award from the Canadian Institutes of Health Research and Canadian Chiropractic Research Foundation. DM is a recipient of a Research Early Career Award from Hamilton Health Sciences Foundation (Jack Hirsh Fellowship). KAOT is funded by unrestricted grants from the Finnish Cultural Foundation and the Finnish Medical Foundation. John You is supported by a Research Early Career Award from Hamilton Health Sciences.

### **Author's contribution**

BK, EvE, and MB designed the study, collected data, interpreted the results and wrote the manuscript. BK and SS managed the database and conducted all analyses which were checked by MB. JY, AB, YK, RS, AA, TB, JJM, MS, KAOT, IN, AC, MF, SMM, DM, EAA, DB, JWB, IF, FL, AN, VG, HR, LM, RR, SE, XS, POV, BCJ, MAW, MS, and LGH contributed to the data collection. BB, HCB, and GHG provided methodological and logistical support. All authors critically revised the manuscript and approved the final version before submission. BK, SS, EvE, and MB are guarantors.

### **Acknowledgements**

We would like to thank the presidents and staff of participating Research Ethics Committees from Switzerland (Basel, Lausanne, Zurich, Lucerne), Germany (Freiburg), and Canada (Hamilton) for their continuous support and cooperation.

### **Ethical approval**

The participating Research Ethics Committees approved the study or explicitly stated that no ethical approval was necessary.

### **Data sharing**

No additional data available.

## References

1. Sun X, Briel M, Busse JW, You JJ, Akl EA, Mejza F, et al. The influence of study characteristics on reporting of subgroup analyses in randomised controlled trials: systematic review. *Bmj*. 2011;342:d1569. PubMed PMID: 21444636. Epub 2011/03/30. eng.
2. Hernandez AV, Boersma E, Murray GD, Habbema JD, Steyerberg EW. Subgroup analyses in therapeutic cardiovascular clinical trials: are most of them misleading? *American heart journal*. 2006 Feb;151(2):257-64. PubMed PMID: 16442886. Epub 2006/01/31. eng.
3. Wang R, Lagakos SW, Ware JH, Hunter DJ, Drazen JM. Statistics in medicine--reporting of subgroup analyses in clinical trials. *The New England journal of medicine*. 2007 Nov 22;357(21):2189-94. PubMed PMID: 18032770. Epub 2007/11/23. eng.
4. Gabler NB, Duan N, Liao D, Elmore JG, Ganiats TG, Kravitz RL. Dealing with heterogeneity of treatment effects: is the literature up to the challenge? *Trials*. 2009;10:43. PubMed PMID: 19545379. Pubmed Central PMCID: 2706823. Epub 2009/06/24. eng.
5. Assmann SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other (mis)uses of baseline data in clinical trials. *Lancet*. 2000 Mar 25;355(9209):1064-9. PubMed PMID: 10744093. Epub 2000/04/01.
6. Szczech LA, Berlin JA, Feldman HI. The effect of antilymphocyte induction therapy on renal allograft survival. A meta-analysis of individual patient-level data. *Anti-Lymphocyte Antibody Induction Therapy Study Group. Annals of internal medicine*. 1998 May 15;128(10):817-26. PubMed PMID: 9599193. Epub 1998/05/23. eng.
7. Rothwell PM, Eliasziw M, Gutnikov SA, Fox AJ, Taylor DW, Mayberg MR, et al. Analysis of pooled data from the randomised controlled trials of endarterectomy for symptomatic carotid stenosis. *Lancet*. 2003 Jan 11;361(9352):107-16. PubMed PMID: 12531577. Epub 2003/01/18. eng.
8. Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. *Annals of internal medicine*. 1992 Jan 1;116(1):78-84. PubMed PMID: 1530753. Epub 1992/01/01.
9. Sun X, Ioannidis JP, Agoritsas T, Alba AC, Guyatt G. How to use a subgroup analysis: users' guide to the medical literature. *JAMA : the journal of the American Medical Association*. 2014 Jan 22-29;311(4):405-11. PubMed PMID: 24449319.
10. Pocock SJ, Assmann SE, Enos LE, Kasten LE. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. *Statistics in medicine*. 2002 Oct 15;21(19):2917-30. PubMed PMID: 12325108. Epub 2002/09/27. eng.
11. Rothwell PM. Treating individuals 2. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. *Lancet*. 2005 Jan 8-14;365(9454):176-86. PubMed PMID: 15639301. Epub 2005/01/11. eng.
12. Yusuf S, Wittes J, Probstfield J, Tyroler HA. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. *JAMA : the journal of the American Medical Association*. 1991 Jul 3;266(1):93-8. PubMed PMID: 2046134. Epub 1991/07/03. eng.
13. Sun X, Briel M, Walter SD, Guyatt GH. Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. *Bmj*. 2010;340:c117. PubMed PMID: 20354011. Epub 2010/04/01.
14. Al-Marzouki S, Roberts I, Evans S, Marshall T. Selective reporting in clinical trials: analysis of trial protocols accepted by *The Lancet*. *Lancet*. 2008 Jul 19;372(9634):201. PubMed PMID: 18640445.
15. Chan AW. Bias, spin, and misreporting: time for full access to trial protocols and results. *PLoS medicine*. 2008 Nov 25;5(11):e230. PubMed PMID: 19067481. Pubmed Central PMCID: 2586359.
16. Chan AW, Hrobjartsson A, Jorgensen KJ, Gotzsche PC, Altman DG. Discrepancies in sample size calculations and data analyses reported in randomised trials: comparison of publications with protocols. *Bmj*. 2008;337:a2299. PubMed PMID: 19056791. Pubmed Central PMCID: 2600604.



17. 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 medical research methodology*. 2012 Aug 28;12(1):131. PubMed PMID: 22928744. Epub 2012/08/30.
18. Kasenda B, von Elm E, You J, Blumle A, Tomonaga Y, Saccilotto R, et al. Prevalence, characteristics, and publication of discontinued randomized trials. *JAMA : the journal of the American Medical Association*. 2014 Mar 12;311(10):1045-51. PubMed PMID: 24618966.
19. Sun X, Briel M, Busse JW, You JJ, Akl EA, Mejza F, et al. Credibility of claims of subgroup effects in randomised controlled trials: systematic review. *Bmj*. 2012;344:e1553. PubMed PMID: 22422832.
20. Chan AW, Upshur R, Singh JA, Ghersi D, Chappuis F, Altman DG. Research protocols: waiving confidentiality for the greater good. *Bmj*. 2006 May 6;332(7549):1086-9. PubMed PMID: 16675819. Pubmed Central PMCID: 1458595. Epub 2006/05/06.
21. Kasenda B, von Elm EB, You J, Blumle A, Tomonaga Y, Saccilotto R, et al. Prevalence, Characteristics, and Publication of Discontinued Randomized Trials. *JAMA : the journal of the American Medical Association*. 2014;[in press].
22. Bhandari M, Devereaux PJ, Li P, Mah D, Lim K, Schunemann HJ, et al. Misuse of baseline comparison tests and subgroup analyses in surgical trials. *Clinical orthopaedics and related research*. 2006 Jun;447:247-51. PubMed PMID: 16672904.
23. Boonacker CW, Hoes AW, van Liere-Visser K, Schilder AG, Rovers MM. A comparison of subgroup analyses in grant applications and publications. *American journal of epidemiology*. 2011 Jul 15;174(2):219-25. PubMed PMID: 21597099. Epub 2011/05/21. eng.
24. ClinicalTrials.gov. Protocol Data Element Definitions (DRAFT) 2013 [cited 2014 March]. Available from: <http://prsinfo.clinicaltrials.gov/definitions.html>.
25. ICTRP. Search Portal 2014 [cited 2014 March]. Available from: <http://www.who.int/ictrp/search/en/>.
26. ISRCTN. Data Set 2012 [cited 2014 March]. Available from: [http://www.controlled-trials.com/isrctn/isrctn\\_dataset](http://www.controlled-trials.com/isrctn/isrctn_dataset).
27. Chan AW, Tetzlaff JM, Altman DG, Laupacis A, Gotzsche PC, Krleza-Jeric K, et al. SPIRIT 2013 statement: defining standard protocol items for clinical trials. *Annals of internal medicine*. 2013 Feb 5;158(3):200-7. PubMed PMID: 23295957. Epub 2013/01/09. Eng.
28. 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. PubMed PMID: 23303884. Pubmed Central PMCID: 3541470. Epub 2013/01/11. eng.
29. Schulz KF, Altman DG, Moher D, Group C. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *Bmj*. 2010;340:c332. PubMed PMID: 20332509. Pubmed Central PMCID: 2844940.

## Tables and Figures

**Table 1:** Trial characteristics based on protocols.

Values are numbers (percentages) unless otherwise specified. IQR, interquartile range

Characteristics	Subgroup analyses		All N=894
	Not planned N=642	Planned N=252	
<b>Target sample size</b>			
Median (IQR)	200 (80, 460)	521 (229, 1007)	260 (100, 606)
<b>Centre status</b>			
Multi centre	500 (77.9)	241 (95.6)	741 (82.9)
Single centre	139 (21.7)	10 (4)	149 (16.7)
Unclear	3 (0.5)	1 (0.4)	4 (0.4)
<b>Study design</b>			
Parallel	592 (92.2)	244 (96.8)	836 (93.5)
Cross-over	40 (6.2)	1 (0.4)	41 (4.6)
Factorial	9 (1.4)	6 (2.4)	15 (1.7)
Unclear	1 (0.2)	1 (0.4)	2 (0.2)
<b>Study intention</b>			
Superiority	456 (71.0)	196 (77.8)	652 (72.9)
Non-inferiority	95 (14.8)	44 (17.5)	139 (15.5)
Unclear	91 (14.2)	12 (4.8)	103 (11.5)
<b>Unit of randomization</b>			
Individuals	629 (98.0)	250 (99.2)	879 (98.3)
Clusters	10 (1.6)	2 (0.8)	12 (1.3)
Body parts	3 (0.5)	0 (0.0)	3 (0.3)
<b>Sponsorship</b>			
Investigator	286 (44.5)	57 (22.6)	343 (39.8)
Industry	356 (55.5)	195 (77.4)	551 (60.2)
<b>Clinical field</b>			
Oncology	113 (17.6)	42 (16.7)	155 (17.3)
Cardiovascular	59 (9.2)	49 (19.4)	108 (12.1)
Infectious disease	60 (9.3)	27 (10.7)	87 (9.7)
Endocrinology	47 (7.3)	15 (6.0)	62 (6.9)
Neurology	37 (5.8)	24 (9.5)	61 (6.8)
Other	326 (50.8)	95 (37.7)	421 (47.1)

**ONLINE TABLE 1:** Characteristics of published trials as reported in journal articles. Values are numbers with percentages unless otherwise specified. IQR, inter quartile range. \* New England Journal of Medicine, Lancet, Journal of the American Medical Association, British Medical Journal, Annals of Internal Medicine, Circulation, Journal of Clinical Oncology, Blood, American Journal of Respiratory and Critical Care Medicine, Paediatrics, Annals of Neurology, Annals of Surgery. + Categorization was based on information from protocol – same as in TABLE 1.

Characteristics	Subgroup analyses		All N=515
	Not reported N=269	Reported N=246	
<b>Sample size</b>			
Median (IQR)	176 (68, 452)	572 (245, 1253)	347 (106, 800)
<b>Centre status</b>			
Multi centre	217 (80.7)	229 (93.1)	446 (86.6)
Single centre	44 (16.4)	12 (4.9)	56 (10.9)
Unclear	8 (3.0)	5 (2.0)	13 (2.5)
<b>Study design</b>			
Parallel	250 (92.9)	232 (94.3)	483 (93.6)
Cross-over	16 (5.9)	7 (2.8)	23 (4.5)
Factorial	2 (0.7)	6 (2.4)	8 (1.6)
Other	1 (0.4)	1 (0.4)	2 (0.4)
<b>Study intention</b>			
Superiority	192 (71.4)	197 (80.1)	389 (75.5)
Non-inferiority	44 (16.4)	40 (16.3)	84 (16.3)
Unclear	33 (12.3)	9 (3.7)	42 (8.2)
<b>Unit of randomization</b>			
Individuals	262 (97.4)	245 (99.6)	507 (98.4)
Clusters	6 (2.2)	1 (0.4)	7 (1.4)
Body parts	1 (0.4)	0 (0)	1 (0.2)
<b>Sponsorship</b> <sup>+</sup>			
Non-industry	99 (36.8)	86 (35.0)	185 (35.9)
Industry	170 (63.2)	160 (65.0)	330 (64.1)
<b>High impact journal</b> <sup>*</sup>			
Yes	52 (19.3)	115 (46.7)	167 (32.4)
No	217 (80.7)	131 (53.3)	348 (67.6)
<b>Clinical field</b>			
Oncology	41 (15.2)	52 (21.1)	93 (18.2)

Cardiovascular	24 (8.9)	51 (20.7)	75 (14.6)
Infectious disease	18 (6.7)	19 (7.7)	37 (7.2)
Neurology	17 (6.3)	18 (7.3)	35 (6.8)
Endocrinology	18 (6.7)	11 (4.5)	29 (5.6)
Other	151 (56.1)	95 (38.6)	246 (47.8)
<b>Results of primary endpoint statistically significant</b>			
Yes	127 (47.2)	122 (49.6)	249 (48.3)
No	101 (37.5)	116 (47.2)	217 (42.1)
No primary endpoint reported	41 (15.2)	8 (3.3)	49 (9.5)

**Table 2:** Subgroup credibility criteria based on trial protocols that planned at least one subgroup analysis, N=252. Values are numbers (percentages) unless otherwise specified. IQR, inter quartile range

Credibility criteria	Sponsorship		All N=252
	Industry N=195	Investigator N=57	
<b>Clear hypothesis given?</b>			
Yes	7 (3.6)	10 (17.5)	17 (6.7)
No	188 (96.4)	47 (82.5)	235 (93.3)
<b>Direction of anticipated effect given?</b>			
Yes	3 (1.5)	7 (12.3)	10 (4.0)
No	192 (98.5)	50 (87.7)	242 (96.0)
<b>Interaction test planned?</b>			
Yes	69 (35.4)	18 (31.6)	87 (34.5)
No	126 (64.6)	39 (68.4)	165 (65.5)
<b>Number of planned subgroup analyses</b>			
Median (IQR)	3 (1, 6)	3 (1, 6)	3 (1, 4)
Not reported (no. of studies)	18 (9.2)	12 (21.1)	30 (11.9)

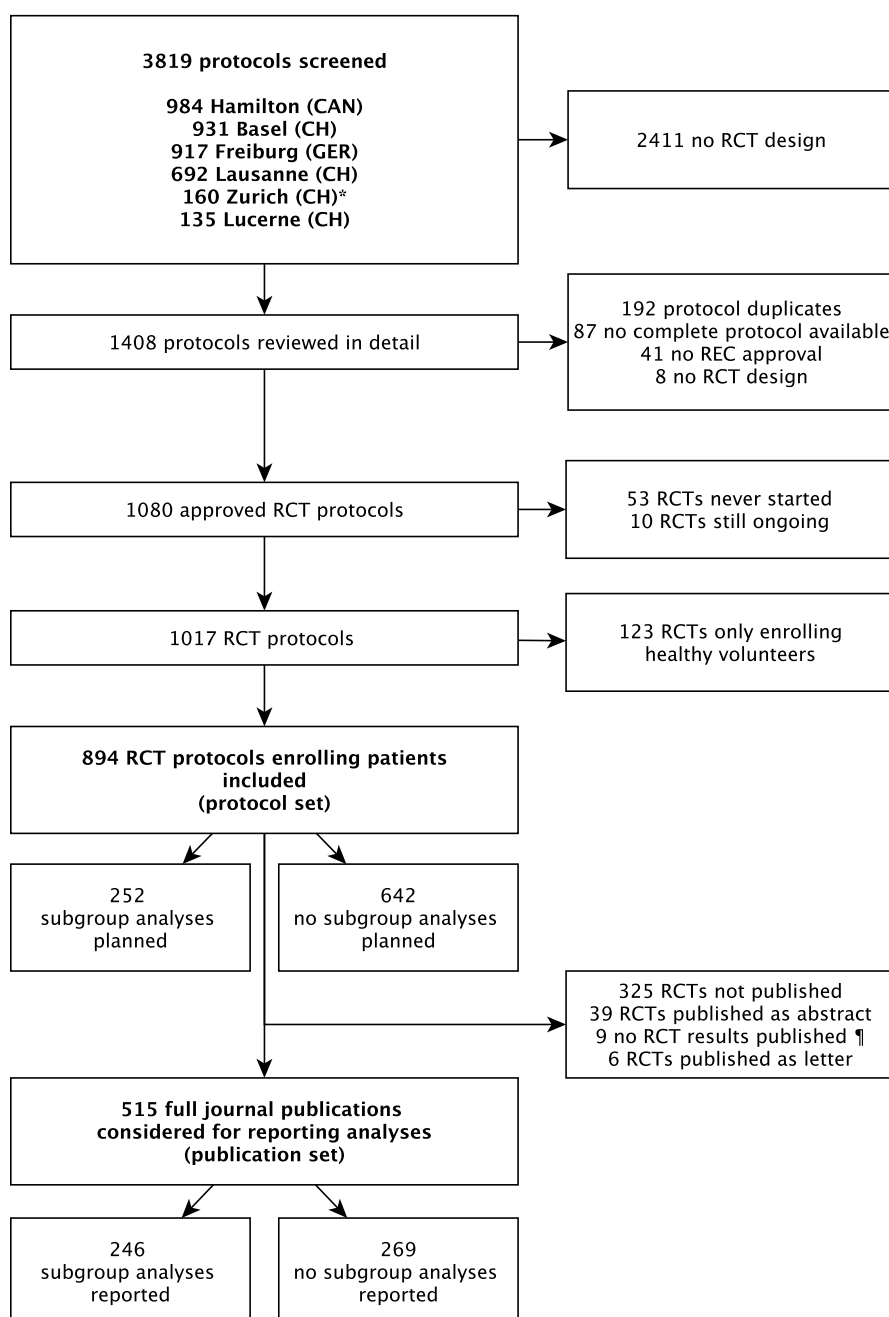
**Table 3:** Reported subgroup credibility criteria and interpretation of subgroup analyses based on publications that reported at least one subgroup analysis, N=246. Values are numbers (percentages) unless otherwise specified. IQR, inter quartile range

Reported credibility criteria and interpretation	Sponsorship		All N=246
	Industry N=160	Investigator N=86	
<b>Pre specification of subgroup analyses reported in publication?</b>			
Yes	58 (36.2)	23 (26.7)	81 (32.9)
No	102 (63.7)	63 (73.3)	165 (67.1)
<b>Post hoc subgroup analyses reported in publication?</b>			
Yes	27 (16.9)	21 (24.4)	48 (19.5)
No	133 (83.1)	65 (75.6)	198 (80.5)
<b>Clear hypothesis given?</b>			
Yes	11 (6.9)	10 (11.6)	21 (8.5)
No	149 (93.1)	76 (88.4)	225 (91.5)
<b>Direction of anticipated effect given?</b>			
Yes	5 (3.1)	6 (7.0)	11 (4.5)
No	155 (96.9)	80 (93.0)	235 (95.5)
<b>Power calculation for subgroup analyses mentioned in publication?</b>			
Yes	3 (1.9)	3 (3.5)	6 (2.4)
No	157 (98.1)	83 (96.5)	240 (97.6)
<b>Test for interaction reported?</b>			
Yes	60 (37.5)	36 (41.9)	96 (39.0)
No	100 (62.5)	50 (58.1)	150 (61.0)
<b>Number of reported subgroup analyses</b>			
Median (IQR)	4 (1, 8)	4 (2, 8)	4 (2, 8)
Not reported	6 (3.8)	2 (1.3)	8 (3.3)
<b>Any claim of subgroup effect reported?</b>			
Yes	57 (35.6)	29 (33.7)	86 (35.0)
No	103 (64.4)	57 (66.3)	160 (65.0)

**Table 4:** Agreement of planning and reporting of subgroup credibility criteria based on those 246 publications reporting at least one subgroup analysis. Numbers are protocols/publications reporting or not reporting subgroup credibility criteria (percentages).

		Reported in publication						
		Industry (N=160)		Investigator (N=86)		All (N=246)		
		No	Yes	No	Yes	No	Yes	
Planned in protocol	Subgroup Hypothesis	No	143 (89.4)	11 (6.9)	73 (84.9)	8 (9.3)	216 (87.8)	19 (7.7)
		Yes	6 (3.8)	0 (0.0)	3 (3.5)	2 (2.3)	9 (3.7)	2 (0.8)
	Direction of effect	No	153 (95.6)	5 (3.1)	78 (90.7)	4 (4.7)	231 (94.0)	9 (3.7)
		Yes	2 (1.3)	0 (0.0)	2 (2.3)	2 (2.3)	4 (1.6)	2 (0.8)
	Interaction test	No	80 (50.0)	46 (28.8)	48 (55.8)	26 (30.2)	128 (52.0)	72 (29.3)
		Yes	20 (12.5)	14 (8.8)	2 (2.3)	10 (11.6)	22 (8.9)	24 (9.8)

**FIGURE 1:** Study flow of RCT protocols and publications. \*In Zurich, we screened only RCT protocols from the two subsidiary Research Ethics Committees responsible for paediatric and surgical RCTs. ¶ No results from randomized comparison published  
Abbreviations: RCT, randomized clinical trial; REC, research ethics committee



### III - Investigation of Continuous Effect Modifiers in a meta-analysis on higher versus lower PEEP in patients requiring mechanical ventilation – protocol of the ICEM study

Benjamin Kasenda, Willi Sauerbrei, Patrick Royston, Matthias Briel

#### Status

**Published**, Kasenda et al. Systematic Reviews 2014, 3:46 (1)

#### Abstract

**Background:** Categorizing an inherently continuous predictor in prognostic analyses raises several critical methodological issues: dependence of the statistical significance on the number and position of the chosen cut-point(s), loss of statistical power, and faulty interpretation of the results if a non-linear association is incorrectly assumed to be linear. This also applies to a therapeutic context where investigators of randomized clinical trials (RCTs) are interested in interactions between treatment assignment and one or more continuous predictors.

**Methods:** Our goal is to apply the multivariable fractional polynomial interaction (MFPI) approach to investigate interactions between continuous patient baseline variables and the allocated treatment in an individual patient data meta-analysis of 3 RCTs (N=2299) from the intensive care field. For each study, MFPI will provide a continuous treatment effect function. Functions from each of the 3 studies will be averaged by a novel meta-analysis approach for functions. We will plot treatment effect functions separately for each study and also the averaged function. The averaged function with a related confidence interval will provide a suitable basis to assess whether a continuous patient characteristic modifies the treatment comparison and may be relevant for clinical decision-making. The compared interventions will be a higher or lower positive end expiratory pressure (PEEP) ventilation strategy in patients requiring mechanical ventilation. The continuous baseline variables body mass index, PaO<sub>2</sub>/FiO<sub>2</sub>, respiratory compliance, and oxygenation index will be the investigated potential effect modifiers. Clinical outcomes for this analysis will be in-hospital mortality, time to death, time to unassisted breathing, and pneumothorax.

**Discussion:** This project will be the first meta-analysis to combine continuous treatment effect functions derived by the MFPI procedure separately in each of several RCTs. Such an approach requires individual patient data (IPD). They are available from an earlier IPD meta-analysis using different methods for analysis. This new analysis strategy allows assessing whether treatment effects interact with continuous baseline patient characteristics and avoids categorisation-based subgroup analyses. These interaction analyses of the present study will be exploratory in nature. However, they may help to foster future research using the MFPI approach to improve interaction analyses of continuous predictors in RCTs and IPD meta-analyses. This study is registered in PROSPERO (CRD42012003129).



## Background

Dichotomizing or categorizing inherently continuous predictor variables raises several issues for statistical analysis and interpretation. These issues include dependence of the statistical significance of the interaction on the number and position of the chosen cut-points, loss of statistical power, and a faulty interpretation of the results if a non-linear association is incorrectly assumed to be linear (2). To overcome these issues, Royston and Sauerbrei proposed the so-called multivariable fractional polynomials interaction (MFPI) approach to investigate potential treatment modifying effects (3, 4). For continuous variables they propose to estimate a treatment effect function, which avoids the well-known problems caused by categorizing continuous variables. To summarize functions across several studies they suggested a new strategy for meta-analysis (5).

A recent individual patient data meta-analysis of 3 randomized controlled trials (RCTs) showed that the pre-defined subgroup of patients who suffered from an acute respiratory distress syndrome had a clinical benefit across various endpoints if they were treated with a higher positive end-expiratory pressure (PEEP) ventilation strategy (6, 7). We will use the MFPI approach (3, 4) and the new strategy to summarize functions across RCTs (5) to re-analyse this dataset of 2299 critically ill patients from the previously reported individual patient data (IPD) meta-analysis (6).

## Objectives

The primary aim of the ICEM study is to demonstrate how methodological issues of interaction/subgroup analyses of continuous predictors can be handled by combining a new meta-analysis approach for functions with the MFPI approach. If IPD are available, MFPI allows investigating whether a continuous variable interacts with treatment in one RCT; combination of data from several RCTs strengthens the assessment concerning a treatment modifying effect. When comparing two (or more) treatments in an RCT, several continuous variables (e.g. age) are suitable candidates to be investigated as potential modifiers of the treatment effect. The ICEM study will be the first example, which combines estimation of treatment effect functions by using MFPI separately in each of several RCTs with a new approach for a meta-analysis of functions. As a secondary aim, we will re-analyse the available IPD data to investigate whether one or more continuous variables have an influence on the comparison of two treatment strategies (higher versus lower PEEP), which is a clinically relevant issue. This paper is an extended version of the registered protocol and shows in an exemplary way how to better use the information from continuous variables if individual patient data from several RCTs is available. In similar projects it should be obvious how to adapt the relevant steps for a meta-analysis of treatment effect functions.

## Methods and Design

Our protocol is registered on PROSPERO (CRD42012003129 at [www.crd.york.ac.uk/prospero/](http://www.crd.york.ac.uk/prospero/)).

### *The dataset*

The present interaction analyses will be based on individual patient data set from 3 RCTs identified by a systematic review in 2010 (6, 8-10) (Table 1, total of 2299 patients). These trials investigated the benefits and harms of higher-PEEP ventilation compared to lower-PEEP ventilation in patients with acute lung injury including acute respiratory distress syndrome. Trial eligibility criteria, literature search strategies, and main results have previously been reported (6). Standardization of variables and consistency checks have already been performed, thus no more data cleaning will be necessary. Before writing the protocol for the study we have updated the earlier (January 2010) literature search (MEDLINE, EMBASE, CENTRAL) and could not identify additional eligible RCTs. Therefore the present analysis will focus on the 3 eligible RCTs from the previous IPD meta-analysis (6).

### *Proposed Statistical Methodology*

#### *Investigation of Interactions*

We will use the MFPI approach (3) to investigate the potential treatment (higher versus lower PEEP) modifying effects of each of the continuous variables with respect to a defined outcome. A 'pair' of a potential modifier (e.g. body mass index [BMI]) and an outcome (e.g. In-hospital mortality) will be considered as one investigation. In total, with 4 potential modifiers and 3 outcomes we will have 12 investigations. There will be no p-value adjustment for multiple investigations. All patients will be analyzed in the group to which they were randomized (intention-to-treat principle). For all analyses we will use the software STATA version 13.0 (Texas, USA).

We will use MFPI with FP2 functions as the most complex allowable function and we will test for an interaction at the 5% level in each trial. FP2 functions are extensions of conventional quadratic functions that provide considerably enhanced flexibility for more realistic modelling in real data. Instead of just powers 1 and 2, they utilize additional combinations of powers of the predictor (see Figure 1 for the powers that may be selected, adapted from Royston and Sauerbrei (2008) (11)). Having just two power terms, FP2 functions can exhibit at most one maximum or minimum. We assume that FP2 functions could be a suitable functional form, assuming that patients with extremely high or low values of the continuous predictor might not benefit from the experimental intervention. For each potential effect modifier the functional relationship between this predictor and the outcome will be illustrated using treatment effect functions, irrespective of the p-value from the test for interaction. The functional form derived with the MFPI procedure will be checked for potential misspecification by considering the treatment effect in 4 subgroups of the predictor of about equal sample size (12). The analysis strategy needs re-consideration if the estimated treatment effect function disagrees severely to the corresponding results in subgroups, indicating misspecification of the treatment effect function. For binary outcomes we will estimate odds ratios with 95% confidence intervals (CIs) to quantify the magnitude of effect. Briel et al. had primarily calculated clinically more intuitive relative risks using log-binomial regression instead of odds ratios, but were confronted with computational problems of non-converging log-binomial models in some analyses. For all binary outcomes they additionally

calculated odds ratios and found similar results although event rates for hospital mortality were >30% in treatment and control groups (Table 6 in (7)). Given the similarity of the results we decided to use the logistic regression model in the present study in order to prevent computational issues when applying the MFPI approach. For survival analysis, Kaplan-Meier estimates and hazard ratios with 95% CIs will be presented. Of note, all investigations of a survival outcome will start with a check of the proportional hazards assumption of the effect of treatment in a univariate Cox-model. We will use the Grambsch-Therneau test for this purpose. If the proportional hazards assumption is seriously violated, we will stop the corresponding investigation and will re-consider a suitable strategy for analysis.

#### *Individual Patient Data Meta-Analysis*

Separately for each study we will conduct an MFPI analysis to estimate a treatment effect function. For each modifier-outcome pair, we will use weighted averaging to obtain a summary treatment effect function based on all 3 studies as previously outlined (5). We will use a fixed effects approach, because we consider 3 studies to be too few for a random effect model although a random distribution can be assumed. Usually this averaged treatment effect function is no FP function. It will be plotted to allow for a qualitative assessment of the possible interaction based on the full information of a potential modifier. The individual functions and the averaged function will be the main results to assess whether the variable is a treatment modifier for the specific outcome. We will not conduct any statistical test for the averaged treatment effect function. Combining p-values from the individual studies would be one possible way to get an overall p-value but this is probably not very helpful. More suitable ways to derive an overall p-value need to be investigated.

#### *Potential Clustering of Data*

We realized that the data of the 3 independent trials are clustered by recruiting hospitals. Although there is evidence of considerable “centre effects” with data from intensive care patients, Briel et al. found that the variance among the 90 recruiting hospitals explained very little (0.3%) of the total variance for hospital mortality (Table 6 in (7)). Differences in patient baseline characteristics such as age, probability of death in hospital from prognostic scores, and proportion of patients with severe sepsis largely (co-variables in the primary analyses of the present study) explained the between-hospital variance of 2.6% found with a basic hierarchical model including only PEEP group and a categorical trial variable as fixed effects and recruiting hospitals as a random effect. Given the negligible between-hospital variance we decided to forgo any consideration of “centre effects” in the primary analyses of the present study.

#### *Adjustment for Confounders*

Because of some imbalances regarding age (8, 9) and the proportion of patients with severe sepsis (9, 10), Briel et al conducted an adjusted analysis for all outcomes (6). We will adopt this approach, thus each analysis will be conducted with adjustment for the following potential confounders: age (continuous), presence of severe sepsis (yes versus no), and predicted probability of dying in the hospital (based on Acute Physiology and Chronic Health Evaluation II and Simplified Acute Physiology II scores, which have similar accuracy (13, 14)).

Selection of these potential confounders resulted from a previous Delphi-like structured survey among experts from the intensive care field (7). We will apply the FP1 function selection strategy to the confounders, with FP1 as the most complex permitted functional form. Including all confounders mentioned above, the model will be determined separately for each of the 3 outcomes using MFP (1.0, 0.05), independent of treatment. In the notation MFP ( $\alpha_1$ ,  $\alpha_2$ ) the value of  $\alpha_1$  gives the significance level for the variable selection part of MFP and  $\alpha_2$  the significance level of the function selection procedure for continuous variables (11). The percentage of missing values among the adjustment variables is much lower than 1%. We will use the singly imputed values (see section “Handling Missing Data”) to replace missing values of these 3 covariates. Despite some imbalances in the covariate distributions between PEEP groups mentioned above, univariate approaches will be conducted as sensitivity analyses.

#### *Influential Points*

To circumvent the issue of influential points all continuous variables will be truncated at the 1% and 99%-tile; meaning that any value below the 1%-tile will be replaced by the value of the 1%-tile, and any value above the 99%-tile will be replaced by the value of the 99%-tile. These truncations will be performed for each study separately.

#### *Handling Missing Data*

Some of the potential modifiers and variables used for adjustment (see below) have missing values of up to about 10%. In order to use all information in all analyses we will impute missing values before the main analysis starts. To try to ensure that the missing at random assumption is valid, we will include all outcomes and as many other variables as possible in the imputation models (15). Five imputations will be created using the multiple imputations by chained equations (MICE) technique. Only the first imputation will be used in analyses. The remaining 4 will be reserved for sensitivity analysis of the main findings.

### **Outcomes**

We selected 3 clinically important outcomes of interest from a larger list of outcomes used in the analysis by Briel et al. (6):

In-hospital mortality at 60 days post randomization (outcome 1a) constitutes the primary efficacy outcome of interest. We will also consider in-hospital mortality as a time-to-event variable (outcome 1b) because we are additionally interested in the timing of mortality events in the randomized groups. Due to the differential follow-up across RCTs beyond day 60 and the fact that the intervention effects happen mainly within the first month, we will censor all surviving patients in the time to event analysis at day 60 as done in the original IPD meta-analysis.

Time to unassisted breathing (outcome 2), which is defined as time from randomization until breathing without mechanical support within the first 28 days is the secondary efficacy outcome of interest. Due to differential follow-up across RCTs for this outcome beyond day 28 and the fact that the intervention effect is supposed to happen before day 28 we will

ensor patients at day 28 as done in the original IPD meta-analysis. Patients who die before achieving unassisted breathing within the first 28 days will be censored at the day of death. With this procedure we circumvent the competing risk issue in the analysis of this outcome. We are aware of the fact that for prognostic questions, which will not be part of this analysis, cumulative incidence functions would be preferred.

Pneumothorax requiring chest tube drainage in first 28 days after randomization (outcome 3, binary variable) is the main safety outcome, because it captures the main potential adverse event directly associated with higher PEEP (experimental intervention). Again, the reason for choosing a 28-day period is that the follow-up for this outcome is different across included trials beyond day 28 and the intervention effect is supposed to happen within the first 28 days. In the present protocol we will not analyze outcome 3 (main safety outcome) because of competing risks with mortality (16). In the planned clinical report of this work we will refer to the results of the original IPD meta-analysis with respect to outcome 3, because the MFPI methodology has still to be adapted for a competing risk framework. We will deal with competing risks in an addition to this protocol. For the specified efficacy outcomes (outcomes 1a/b and 2) we anticipate no competing risk problems when using cause-specific Cox models.

The following 4 continuous potential effect modifiers were all pre-specified by Briel et al (6):

#### Body Mass Index (BMI) at baseline

The BMI is calculated by the ratio of mass in  $\text{kg}/\text{m}^2$ . There is no data that suggest a certain direction of the treatment effect modification, but Briel et al hypothesized less benefit of higher PEEP in patients with higher BMI (6).

#### Respiratory Compliance (RC) at baseline

The RC is estimated by the ratio of the tidal volume in ml / (inspiratory plateau pressure-PEEP in  $\text{cm H}_2\text{O}$ ). A lower RC would reflect more severe lung injury. Briel et al. hypothesized that patients with lower RC have more recruitable lung units and would therefore benefit from higher levels of PEEP. In addition, one could argue that in patients suffering from most severe ARDS, which is commonly associated with very low RC, higher PEEP might no longer provide any benefit.

#### $\text{PaO}_2/\text{FiO}_2$ ratio at baseline

A low  $\text{PaO}_2/\text{FiO}_2$  reflects impaired blood oxygenation and therefore more severe lung injury. Similar to RC Briel et al. hypothesized that patients with a lower  $\text{PaO}_2/\text{FiO}_2$  ratio benefit more from higher PEEP levels. It will be interesting to see how the widely accepted ARDS defining cut-off at 200mmHg is reflected in this analysis using the MFPI approach. Using this cut-off, a significant interaction was found by Briel et al (6, 7).

#### Oxygenation Index at baseline

The oxygenation index (OI, defined as mean airway pressure x 100 / [ $\text{PaO}_2/\text{FiO}_2$  ratio]) includes the mean airway pressure and can be regarded as the more reliable marker regarding blood oxygenation compared to the  $\text{PaO}_2/\text{FiO}_2$  ratio alone. The higher the OI, the

more severe the lung injury; therefore Briel et al. hypothesized that patients with a higher OI benefit more from higher PEEP levels (6).

Further candidates (e.g age and sex) may be additionally investigated for interaction. Of note, irrespective of the results, all investigations will be included in a summary table similar to the REMARK profile for prognostic studies (17).

## Discussion

The ICEM study is the first example, which combines estimation of treatment effect functions by using MFPI with a new approach for a meta-analysis of functions for a clinically relevant issue. The approach requires IPD data, which are available from an earlier meta-analysis project. The present article is an extended version of the registered protocol and shows in an exemplary way how to better use the information from continuous variables if individual patient data from several RCTs are available. In similar projects it should be obvious how to adapt the relevant steps for a meta-analysis of treatment effect functions. Besides the new application of the MFPI approach in meta-analysis, the available dataset from 3 RCTs also offers a unique opportunity to identify potential clinically important interaction effects. All these interaction analyses are exploratory in nature; however, they use the full information for a potential treatment modifier and may help in clinical decision-making. We hope that this project will also foster future research using the MFPI approach to improve interaction analyses of continuous predictors in RCTs and in meta-analyses, provided individual patient data is available.

## Competing interests and funding

This project has no specific funding. MB is supported by santésuisse and the Gottfried and Julia Bangerter-Rhyner Foundation. The funding sources have no role in the design and conduct of this study and the writing of this manuscript. The article processing charge was funded by the German Research Foundation (DFG) and the Albert Ludwigs University Freiburg in the funding program Open Access Publishing.

## Authors' contribution

BK, WS, PR, and MB have designed the study and written the registered protocol and this manuscript. WS and PR developed the multivariable fractional polynomials interaction (MFPI) and the meta-analysis approach to combine several functions across studies. MB provided the database. BK, WS, and PR will conduct the analyses. All authors approved the final version before submission.

## Acknowledgements

We would like to thank all investigators from the original randomized controlled trials included in this individual patient data meta-analysis (namely Maureen Meade, Alain Mercat, Roy G. Brower, Arthur S. Slutsky, Deborah Cook, Gordon H. Guyatt, Laurent

Brochard, Jean-Christophe M. Richard, and Thomas E. Stewart) for their continuous support and cooperation.

## References

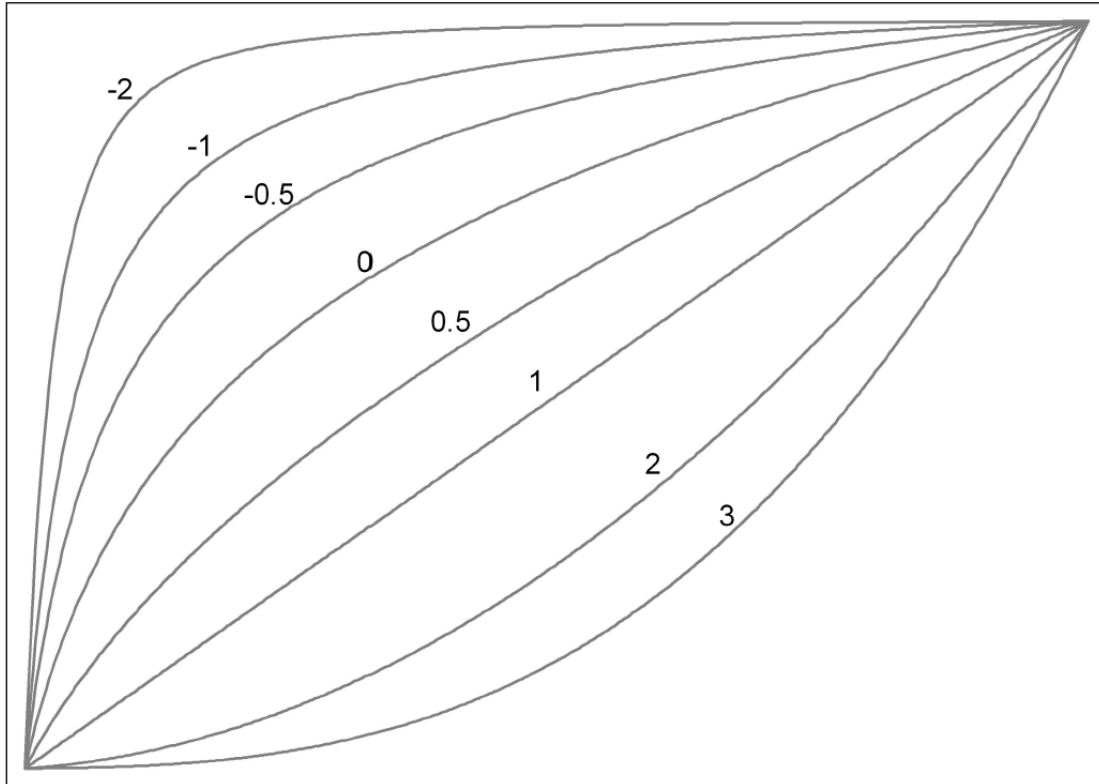
1. Kasenda B, Sauerbrei W, Royston P, Briel M. Investigation of continuous effect modifiers in a meta-analysis on higher versus lower PEEP in patients requiring mechanical ventilation - protocol of the ICEM study. *Systematic reviews*. 2014;3(1).
2. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Statistics in medicine*. 2006 Jan 15;25(1):127-41. PubMed PMID: 16217841.
3. Royston P, Sauerbrei W. A new approach to modelling interactions between treatment and continuous covariates in clinical trials by using fractional polynomials. *Statistics in medicine*. 2004 Aug 30;23(16):2509-25. PubMed PMID: 15287081. Epub 2004/08/03.
4. Royston P, Sauerbrei W. Interactions between treatment and continuous covariates: a step toward individualizing therapy. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2008 Mar 20;26(9):1397-9. PubMed PMID: 18349388.
5. Sauerbrei W, Royston P. A new strategy for meta-analysis of continuous covariates in observational studies. *Statistics in medicine*. 2011 Dec 10;30(28):3341-60. PubMed PMID: 21953493. Epub 2011/09/29.
6. Briel M, Meade M, Mercat A, Brower RG, Talmor D, Walter SD, et al. Higher vs lower positive end-expiratory pressure in patients with acute lung injury and acute respiratory distress syndrome: systematic review and meta-analysis. *JAMA : the journal of the American Medical Association*. 2010 Mar 3;303(9):865-73. PubMed PMID: 20197533. Epub 2010/03/04.
7. Briel M. Higher versus lower positive end-expiratory pressure in patients with acute lung injury: a systematic review and individual patient data meta-analysis (study protocol) [Master Thesis]. Hamilton, Canada: McMaster University; 2008.
8. Brower RG, Lanken PN, MacIntyre N, Matthay MA, Morris A, Ancukiewicz M, et al. Higher versus lower positive end-expiratory pressures in patients with the acute respiratory distress syndrome. *The New England journal of medicine*. 2004 Jul 22;351(4):327-36. PubMed PMID: 15269312. Epub 2004/07/23.
9. Meade MO, Cook DJ, Guyatt GH, Slutsky AS, Arabi YM, Cooper DJ, et al. Ventilation strategy using low tidal volumes, recruitment maneuvers, and high positive end-expiratory pressure for acute lung injury and acute respiratory distress syndrome: a randomized controlled trial. *JAMA : the journal of the American Medical Association*. 2008 Feb 13;299(6):637-45. PubMed PMID: 18270352. Epub 2008/02/14.
10. Mercat A, Richard JC, Vielle B, Jaber S, Osman D, Diehl JL, et al. Positive end-expiratory pressure setting in adults with acute lung injury and acute respiratory distress syndrome: a randomized controlled trial. *JAMA : the journal of the American Medical Association*. 2008 Feb 13;299(6):646-55. PubMed PMID: 18270353. Epub 2008/02/14.
11. Royston P, Sauerbrei W. *Multivariable model-building. A pragmatic approach to regression analysis based on fractional polynomials for modelling continuous variables*. Chichester: JohnWiley & Sons; 2008.
12. Royston P, Sauerbrei W, Ritchie A. Is treatment with interferon-alpha effective in all patients with metastatic renal carcinoma? A new approach to the investigation of interactions. *British journal of cancer*. 2004 Feb 23;90(4):794-9. PubMed PMID: 14970855. Pubmed Central PMCID: 2410187.
13. Beck DH, Smith GB, Pappachan JV, Millar B. External validation of the SAPS II, APACHE II and APACHE III prognostic models in South England: a multicentre study. *Intensive care medicine*. 2003 Feb;29(2):249-56. PubMed PMID: 12536271. Epub 2003/01/22.

14. Suistomaa M, Niskanen M, Kari A, Hynynen M, Takala J. Customized prediction models based on APACHE II and SAPS II scores in patients with prolonged length of stay in the ICU. *Intensive care medicine*. 2002 Apr;28(4):479-85. PubMed PMID: 11967604. Epub 2002/04/23.
15. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in medicine*. 2011 Feb 20;30(4):377-99. PubMed PMID: 21225900. Epub 2011/01/13.
16. Wolkewitz M, Beyersmann J, Gastmeier P, Schumacher M. Modeling the effect of time-dependent exposure on intensive care unit mortality. *Intensive care medicine*. 2009 May;35(5):826-32. PubMed PMID: 19183943. Epub 2009/02/03.
17. Altman DG, McShane LM, Sauerbrei W, Taube SE. Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK): explanation and elaboration. *PLoS medicine*. 2012;9(5):e1001216. PubMed PMID: 22675273. Pubmed Central PMCID: 3362085.



## Figures

**FIGURE 1:** The variety of curve shapes available with the FP1 family of transformations of a continuous predictor,  $x$ . FP1 transformations are simply powers of the form  $x^p$ . For example,  $x^p$  with power  $p = -1$  is the reciprocal ( $1/x$ ) of  $x$ . These powers are indicated by the numbers on the diagram. Adapted from Royston and Sauerbrei (2008), with permission from John Wiley and Sons Ltd. [18].



## IV - Investigation of Continuous Effect Modifiers in a meta-analysis on higher versus lower PEEP in ventilated patients with ARDS – the ICEM study

Benjamin Kasenda, Willi Sauerbrei, Patrick Royston, Alain Mercat, Roy G. Brower, Arthur S. Slutsky, Deborah Cook, Gordon H. Guyatt, Laurent Brochard, Jean-Christophe M. Richard, Thomas E. Stewart, Maureen Meade, and Matthias Briel

### Status

**First manuscript draft finished.** Planned to be submitted by June 2014.

### Abstract

**Background.** Low tidal volumes and low inspiratory pressures are considered standard to prevent lung injury, but the optimal level of positive end-expiratory pressure (PEEP) is still unclear. A recent individual patient (IPD) data meta-analysis suggested that patients with moderate or severe acute respiratory distress syndrome (ARDS) benefit from high PEEP ventilation strategies.

**Objectives.** To investigate interactions between continuous patient baseline variables and higher versus lower PEEP using the multivariable fractional polynomial interaction (MFPI) approach.

**Patients and Methods.** We re-analysed IPD of 2299 ventilated patients from 3 randomized trials identified by a systematic review that investigated benefits and harms of higher (N=1136) versus lower PEEP (N=1163) ventilation strategies. We examined the following continuous baseline characteristics: body mass index [ $\text{kg}/\text{m}^2$ ],  $\text{PaO}_2/\text{FiO}_2$  [ratio of partial pressure arterial  $\text{O}_2$  and fraction of inspired  $\text{O}_2$ ], respiratory compliance [ $\text{ml}/\text{cm H}_2\text{O}$ ], and oxygenation index [mean airway pressure  $\times 100 / (\text{PaO}_2/\text{FiO}_2 \text{ ratio})$ ]. Outcome measures were in-hospital mortality, time to death, and time to unassisted breathing. For each trial, MFPI provided a continuous treatment effect function (TEF). Functions from each of the 3 trials were averaged by a novel meta-analysis approach for functions. We investigated interaction using the plotted summary TEF by qualitative assessment.

**Results.** The summary TEFs for  $\text{PaO}_2/\text{FiO}_2$  revealed a U-shaped curve suggesting that mostly patients with values between 150 mmHg (Odds ratio [OR] 0.85, 95% CI 0.62 - 1.03) and 100 mmHg (OR 0.85, 95% CI 0.65 - 1.07) benefit from high PEEP ventilation strategies with respect to all three outcomes. Patients with more extreme values at both ends did not seem to benefit from higher PEEP levels and could even be harmed. Patients with respiratory compliance values above 40  $\text{ml}/\text{cm H}_2\text{O}$  (OR 0.80, 95% CI 0.61 - 1.04) showed a steadily growing benefit from higher PEEP levels with respect to mortality and time to death. We found some evidence that patients with higher body mass index (above 35  $\text{kg}/\text{m}^2$ ) may benefit from higher PEEP ventilation strategies with respect to 60 days in-hospital mortality and time to death.

**Conclusions.** Patients with  $\text{PaO}_2/\text{FiO}_2$  between 100 and 150mmHg benefit most from high PEEP ventilation strategies with respect to 60 days in-hospital mortality, time to death, and time to unassisted breathing. Also patients with a good respiratory compliance showed a benefit from higher PEEP levels.

## Background

Acute respiratory distress syndrome (ARDS) carries a high mortality rate of over 40% (1, 2). According to guidelines of the American European Consensus Conference, a patient suffers from ARDS if the ratio of the arterial partial oxygen pressure ( $\text{PaO}_2$ ) to the fraction of inspired oxygen ( $\text{FiO}_2$ ) is below 200mmHg (3). Most recently, the ARDS definition task force proposed three stages of ARDS severity based on degree of hypoxemia: mild ( $200\text{mmHg} < \text{PaO}_2/\text{FiO}_2 \leq 300 \text{ mmHg}$ ), moderate ( $100\text{mmHg} < \text{PaO}_2/\text{FiO}_2 \leq 200\text{mmHg}$ ), and severe ( $\text{PaO}_2/\text{FiO}_2 < 100\text{mmHg}$ ) (4). Low tidal volumes and low inspiratory pressures are considered standard to prevent lung injury (5, 6), but the optimal level of positive end-expiratory pressure (PEEP) is still under debate. Recently, Briel et al conducted a systematic review and individual patient data (IPD) meta-analysis of three randomized clinical trials (RCTs) to investigate higher versus lower PEEP ventilation in critically ill patients. They concluded that higher-PEEP ventilation strategies were not superior compared to lower PEEP levels; however, within the pre-defined subgroup of patients who suffered from moderate or severe ARDS, a benefit was suggested across various clinical outcome measures including time to un-assisted breathing and overall hospital mortality (7, 8). Briel et al. used the widely accepted 200mmHg cut-off (3) to examine this subgroup of ARDS patients using state of the art statistical methods including tests for interaction. They also conducted interaction analyses of other continuous variables e.g. body mass index by assuming linearity or categorizing them into quintiles. It is well recognized that dichotomizing or categorizing an inherently continuous predictor (e.g.  $\text{PaO}_2/\text{FiO}_2$ , age, or body mass index) may raise several critical methodological issues: dependence of the statistical significance on the number and position of the chosen cut-points, loss of statistical power, and faulty interpretation of the results if a non-linear association is incorrectly assumed to be linear (9). To overcome these weaknesses, Royston and Sauerbrei proposed the multivariable fractional polynomials interaction (MFPI) approach to investigate potential treatment modifying effects (10-12) in a RCT. MFPI estimates a continuous function to quantify the relative effect of two treatments depending on a continuous predictor of interest (treatment effect function, TEF). To derive a summary estimate of functional relationship across several studies, Sauerbrei and Royston proposed a new strategy, which they derived from the content of meta-analysis of a continuous prognostic factor (12). To investigate potential treatment interactions of continuous predictor variables in their inherent form, we re-analysed the previously reported individual patient data meta-analysis (7) with respect to in-hospital mortality, time-to-death, and time-to-unassisted breathing using the MFPI approach.

## Patients and Methods

The rationale and methodological background of this study have been described elsewhere (13); the protocol is registered at PROSPERO (CRD42012003129 at <http://www.crd.york.ac.uk/prospero>). Briefly, the present analyses were based on IPD from three RCTs (14-16) (N=2299 patients) all identified through a systematic review of the literature (7). All RCTs investigated the benefits and harms of higher-PEEP ventilation compared to lower-PEEP ventilation in patients with ARDS. Inclusion criteria, literature search strategies, and main results of this IPD meta-analysis have been reported previously; FIGURE 1 illustrates the trial flow (7).

*Clinical outcomes*

In-hospital mortality at 60 days post randomization (outcome 1a) constituted the primary outcome of interest. We also considered in-hospital mortality as a time-to-event variable (outcome 1b) because we were additionally interested in the timing of mortality events in the randomized groups. All trials (LOVS, ALVEOLI, EXPRESS) followed-up patients for at least 60 days; thereafter LOVS and ALVEOLI followed-up the majority of patients until death or hospital discharge but EXPRESS did not (7). Due to the differential follow-up across RCTs beyond day 60 and the fact that the intervention effects happen mainly within the first month, we censored all surviving patients in the time to event analysis at day 60 as done in the original IPD meta-analysis (7). We also investigated time to unassisted breathing (outcome 2), which was defined as time from randomization until breathing without mechanical support within the first 28 days. Because of differential follow-up across RCTs for this outcome beyond day 28 and the fact that the intervention effect is supposed to happen before day 28 we censored patients at day 28 as done in the original IPD meta-analysis (7). Patients who died before achieving unassisted breathing within the first 28 days were censored at the day of death. With this procedure we have circumvented the competing risk issue in the analysis of this outcome. As described in our protocol, pneumothorax requiring chest tube drainage (outcome 3, binary variable) during the first 28 days after randomization is the main safety outcome, and supposed to capture the main potential adverse effects directly associated with higher PEEP (experimental intervention). The MFPI methodology has still to be adapted for a competing risk framework, therefore we have not analysed this safety outcome, because of competing risks with mortality (17).

*Potential effect modifiers*

Body mass index at baseline: The body mass index was calculated by the ratio of body weight and height in metres squared ( $\text{kg}/\text{m}^2$ ). The current analysis is exploratory in nature, but Briel et al hypothesized less benefit of higher PEEP in patients with higher body mass index (7).

Respiratory compliance at baseline: The respiratory compliance was estimated by the ratio of the tidal volume in ml divided by the inspiratory plateau pressure PEPP in  $\text{cm H}_2\text{O}$  ( $\text{ml}/\text{cm H}_2\text{O}$ ). A lower respiratory compliance would reflect more severe lung injury. Briel et al. hypothesized that patients with lower respiratory compliance have more recruitable lung units and would therefore benefit from higher levels of PEEP (7).

$\text{PaO}_2/\text{FiO}_2$  at baseline: A low  $\text{PaO}_2/\text{FiO}_2$  reflects impaired blood oxygenation and therefore more severe lung injury. Similar to respiratory compliance, Briel et al. hypothesized that patients with a  $\text{PaO}_2/\text{FiO}_2$  ratio  $< 200$  mmHg benefit more from higher PEEP levels. Using this cut-off, a significant interaction was found by Briel et al (7, 8).

Oxygenation index at baseline: The oxygenation index (defined as mean airway pressure times  $100 / [\text{PaO}_2/\text{FiO}_2]$ ) includes the mean airway pressure and can be regarded as the more reliable marker regarding blood oxygenation compared to the  $\text{PaO}_2/\text{FiO}_2$  alone. The higher the oxygenation index, the more severe the lung injury; therefore Briel et al.

hypothesized that patients with a higher oxygenation index benefit more from higher PEEP levels (7).

## Methods for analysis

### *Multivariable fractional polynomial interaction (MFPI) procedure*

For each individual study, we used the MFPI approach (10) to investigate potential interactions between treatment assignment (higher versus lower PEEP) and one of the four potential modifiers with respect to mortality, time to death, and time to unassisted breathing. We considered fractional polynomial 2 (FP2) functions as the most complex allowable function; two power terms giving the best fit in both treatment groups were determined. To quantify the magnitude of effects, we estimated odds or hazard ratios with 95% confidence intervals (CI) as a continuous function. We calculated p-values of the test for an interaction and plotted TEFs separately for each study.

### *Averaging the functions of individual studies*

We conducted an MFPI analysis with all included patients stratified by trial. For each modifier-outcome pair, we used weighted averaging (fixed-effect) to obtain a summary TEF based on all 3 RCTs as previously outlined (12). These averaged TEFs were plotted to allow for a qualitative assessment of the possible interaction. The fixed-effects weights for averaging the functions across studies were derived from the reciprocal of the variances; point wise 95% CIs were calculated accordingly (12). For each potential modifier we illustrated the study weights graphically. Although it is possible to calculate a combined p-value for the meta-analysis of several functions, we did not calculate it because treatment effect functions were non-linear and the qualitative assessment of the functional relationship is of central interest and not a single test for significance.

### *Adjustment for confounders*

All patients were analysed in the group to which they were randomized (intention-to-treat principle). Because of some imbalances with respect to age (14, 15) and the proportion of patients with severe sepsis (15, 16), all MFPI analyses were adjusted for the following potential confounders: age (continuous), presence of severe sepsis (yes versus no), and predicted probability of dying in the hospital (based on Acute Physiology and Chronic Health Evaluation II and Simplified Acute Physiology II scores, which have similar accuracy (18, 19)). We applied the FP1 function selection strategy to the confounders, with FP1 as the most complex permitted functional form.

### *Missing values and influential points*

Some of the potential modifiers and variables used for adjustment had missing values of up to about 30%. In order to use all information in all analyses we imputed missing values by multiple imputations chained equations techniques (20). To circumvent the issue of influential outliers we truncated each continuous predictor at the 1% and 99% percentile; meaning that values below / above the 1% / 99% percentile were replaced by the value of

the 1% / 99% percentile, respectively. For all analyses we used the software STATA version 13.0 (Texas, USA).

## Results

### *Patient characteristics*

TABLE 1 summarizes patients' characteristics and clinical outcomes stratified by trial. In the LOVS trial, more patients were allocated to the lower-PEEP group, because of initial problems due to a programming error of the blocked central randomization. However, sensitivity analyses conducted by the investigators showed consistent results across various assumptions in this trial (15). FIGURE 1 illustrates the trial flow as previously reported (7).

FIGURE 2 illustrates the adjusted summary TEFs for each modifier-outcome pair. The distributions of weights given to each trial to calculate the averaged TEFs are summarized in FIGURE 3. The shape of these curves is a result of the distribution of events by the respective modifier. For example, whereas the EXPRESS trial has more mortality events (280 versus 134) compared to the ALVEOLI trial in the group of patients with body mass index less than 40 kg/m<sup>2</sup>, the ALVEOLI trial contributed more events (10 versus 5) in the group of patients with body mass index above 40 kg/m<sup>2</sup> (FIGURE 3, left upper cell). Because we mainly focused on the meta-analysis in this manuscript, we provide the single TEFs for each RCT and modifier-outcome pair in the ONLINE APPENDIX FIGURES 1 - 3.

### *Interaction with body mass index*

The first row of FIGURE 2 shows TEFs averaged over the three individual RCTs illustrating the interaction between body mass index and the clinical outcomes. For 60 days in-hospital mortality, the shape of the curve suggests a trend that patients with higher BMI may benefit from higher PEEP levels. However, the upper 95% CI always includes one and the uncertainty of the estimated effect grows rapidly once BMI is above 40 kg/m<sup>2</sup>. For the interaction with the outcome time to death, the curve shows a very similar pattern. For time to unassisted breathing, no interaction can be assumed, because the estimated line is almost parallel to the x-axis. TEFs of the individual trials differ slightly and not any of the individual RCTs showed a significant interaction between body mass index and PEEP intervention for any of the three outcomes (APPENDIX FIGURES 1 - 3).

### *Interaction with respiratory compliance*

The second row of FIGURE 2 shows TEFs averaged over the three individual RCTs illustrating the interaction between respiratory compliance and the clinical outcomes. Regarding 60 days in-hospital mortality and time to death, the monotonically decreasing TEF curves of both outcomes suggest that patients with better respiratory compliance benefit from higher PEEP levels. After the value of 40 ml/cm H<sub>2</sub>O (OR for in-hospital mortality 0.80, 95% CI 0.61 - 1.04), the upper 95% CI limit almost excludes the OR of one, which also provides more certainty about this positive interaction. TEFs of the individual trials differ slightly and not any of the individual RCTs showed a significant interaction between respiratory compliance and PEEP intervention for any of the three outcomes (APPENDIX FIGURES 1 - 3).

*Interaction with PaO<sub>2</sub>/FiO<sub>2</sub>*

The third row of FIGURE 2 shows TEFs averaged over the three individual RCTs illustrating the interaction between PaO<sub>2</sub>/FiO<sub>2</sub> and the outcomes. For 60 days in-hospital mortality and time to death, the U-shaped curves suggest that patients with a PaO<sub>2</sub>/FiO<sub>2</sub> between 100 mmHg (OR 0.85, 95% CI 0.65 - 1.07) and 150 mmHg (OR 0.85, 95% CI 0.62 - 1.03) benefit from higher PEEP levels, this is in particular pronounced in the TEF for time to death, e.g. at 125 mmHg (HR 0.82, 95% CI 0.69 – 0.98). The ORs for patients with values above 150mmHg steadily increase and cross the OR of 1 at 200mmHg (OR 1.00, 95% CI 0.72 - 1.39). Regarding time to unassisted breathing, the TEF somehow mirrors the positive interaction, suggesting that patients with PaO<sub>2</sub>/FiO<sub>2</sub> between 100 and 150 mmHg have a relative shorter time to unassisted breathing when treated with higher PEEP values. Especially the TEF of the LOVS trial is also U-shaped similar to the averaged TEF (APPENDIX FIGURE 1).

*Interaction with oxygenation index*

The fourth row of FIGURE 2 shows TEFs averaged over the three individual RCTs illustrating the interaction between oxygenation index and the clinical outcomes. Regarding 60 days in-hospital mortality and time to death, the rather flat shape of both TEFs do not provide much evidence for interaction. Moreover, no significant interaction was observed with PEEP assignment in any of the individual RCTs (APPDENDIX FIGURES 1 - 3).

**Discussion***Summary of findings*

Patients with PaO<sub>2</sub>/FiO<sub>2</sub> values between 100 and 150mmHg may benefit from higher PEEP ventilation strategies with respect to 60 days in-hospital mortality, time to death, and time to unassisted breathing. We found a potential interaction between PEEP level and respiratory compliance above 40 ml/cm H<sub>2</sub>O. There was some evidence, that patients with higher body mass index (above 35 kg/m<sup>2</sup>) may benefit from higher PEEP ventilation strategies with respect to 60 days in-hospital mortality and time to death; however, the uncertainty around this positive interaction is high because of few very obese patients in this analysis set.

*Comparison to the original analysis*

Our primary analysis was based on the imputed dataset including 2299 patients. In the original analysis, Briel et al considered the complete case data set for their primary analysis and conducted one sensitivity analysis based on an imputed dataset, which did not change the conclusion of the primary analysis. In the original analysis, high-PEEP ventilation strategies improved outcome in patients with moderate and severe ARDS at baseline as defined by a PaO<sub>2</sub>/FiO<sub>2</sub> ratio < 200mmHg (7). However, when Briel et al investigated the interaction using the continuous variable assuming linearity, no significant interaction was observed. Of note, Briel et al did not investigate the possible interaction graphically as we did using the plot of the TEF. Therefore, Briel et al reasoned about a possible threshold effect rather than a continuous interaction. At the time of their analysis, the new ARDS

categorization (mild 200 - 300 mmHg PaO<sub>2</sub>/FiO<sub>2</sub>, moderate 100 - 200 mmHg PaO<sub>2</sub>/FiO<sub>2</sub>, and severe <100mmHg PaO<sub>2</sub>/FiO<sub>2</sub>) was not defined yet (4). Considering these proposed cut-points, the qualitative assessment of the TEF for PaO<sub>2</sub>/FiO<sub>2</sub> from our MFPI analysis suggests that patients with moderate ARDS may benefit most from higher PEEP ventilation strategies, and that benefit for patients with mild or severe ARDS is questionable. In summary, the results from our MFPI analysis are not contradictory to the subgroup effect initially identified by Briel et al, but they provide more information and allow a qualitative assessment of the interaction using the resulting plot of the TEF adjusted for potential confounders. This is a critical issue, because relying on statistical significance without any qualitative investigation of the interaction may discard important information that could be useful in clinical decision-making and management, and to generate new hypothesis that could be specifically tested in future trials.

In our analysis, the averaged TEF of BMI showed some trend that patients with higher BMI might benefit more from high-PEEP ventilation strategies compared to patients with lower BMI. Interestingly, Briel et al initially hypothesized that higher PEEP in patients with higher BMI would not improve outcome, because of fewer recruitable lungs units (7). Very obese patients were not included in the three trials, therefore, although the evidence for interaction in the present analysis is rather weak, we assume that if more obese patients would have been included, the beneficial effect for very obese patients would have become more clear.

#### *Strengths and limitations*

We used all available information from the whole dataset and did not depend on any cut-points for our interaction analyses, whether suitably chosen or not. We hereby maximized the statistical power and allowed for non-linear associations, which turned out to describe identified interactions well. A simulation study to investigate power issues when using the MFPI approach is currently under revision. In addition, the strengths of this analysis include an explicit study protocol and analysis plan in which we comprehensively described the clinical variables to be investigated for interaction with the assigned intervention; the study protocol was registered online and is freely accessible.

There is one additional eligible ARDS trial, which has recently completed recruitment of 224 patients (NCT00431158). Although desirable, IPD from this trial could not be included in our dataset so far. After publication of this trial, we will make any effort to include these IPD in our meta-analysis to further increase the precision of our interaction estimates.

#### *Implications and Conclusions*

To our knowledge, this is the first study that used the MFPI approach to investigate possible interactions between continuous clinical predictors and treatment assignments in the framework of an IPD meta-analysis. Our results do not allow for definite conclusions regarding actual clinical care, however, we suggest that the possible benefit for higher PEEP ventilation strategies for patients with moderate ARDS should be considered in their management. Furthermore, the potential interaction between BMI and high PEEP levels should be addressed in future ICU trials, because the incidence of obesity is increasing (21) and the obese population might benefit from tailored ventilation strategies for which evidence is still sparse.



### Conclusions

This IPD meta-analysis suggests that ventilated patients with PaO<sub>2</sub>/FiO<sub>2</sub> between 100 and 150mmHg might benefit most from higher PEEP ventilation strategies with respect to 60 days in-hospital mortality, time-to-death, and time-to-unassisted breathing. Also patients with a respiratory compliance above 40 ml/cm H<sub>2</sub>O may derive a benefit from higher PEEP levels. If IPD are available, the MFPI is a straightforward method to investigate interactions between continuous predictors and outcomes by using all information available in the dataset.

### References

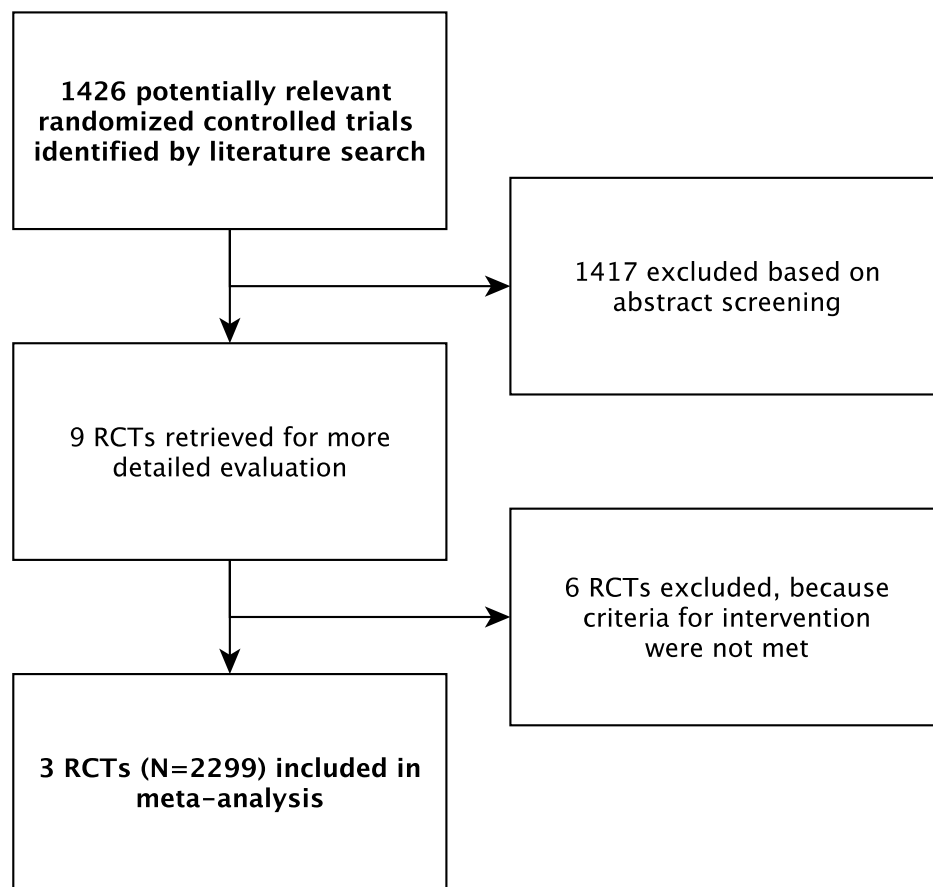
1. Villar J, Blanco J, Anon JM, Santos-Bouza A, Blanch L, Ambros A, et al. The ALIEN study: incidence and outcome of acute respiratory distress syndrome in the era of lung protective ventilation. *Intensive care medicine*. 2011 Dec;37(12):1932-41. PubMed PMID: 21997128.
2. Villar J, Sulemanji D, Kacmarek RM. The acute respiratory distress syndrome: incidence and mortality, has it changed? *Current opinion in critical care*. 2014 Feb;20(1):3-9. PubMed PMID: 24309954.
3. Bernard GR, Artigas A, Brigham KL, Carlet J, Falke K, Hudson L, et al. The American-European Consensus Conference on ARDS. Definitions, mechanisms, relevant outcomes, and clinical trial coordination. *American journal of respiratory and critical care medicine*. 1994 Mar;149(3 Pt 1):818-24. PubMed PMID: 7509706.
4. Force ADT, Ranieri VM, Rubenfeld GD, Thompson BT, Ferguson ND, Caldwell E, et al. Acute respiratory distress syndrome: the Berlin Definition. *JAMA : the journal of the American Medical Association*. 2012 Jun 20;307(23):2526-33. PubMed PMID: 22797452. Epub 2012/07/17.
5. Gattinoni L, Caironi P. Refining ventilatory treatment for acute lung injury and acute respiratory distress syndrome. *JAMA : the journal of the American Medical Association*. 2008 Feb 13;299(6):691-3. PubMed PMID: 18270359.
6. Dellinger RP, Levy MM, Rhodes A, Annane D, Gerlach H, Opal SM, et al. Surviving Sepsis Campaign: international guidelines for management of severe sepsis and septic shock, 2012. *Intensive care medicine*. 2013 Feb;39(2):165-228. PubMed PMID: 23361625.
7. Briel M, Meade M, Mercat A, Brower RG, Talmor D, Walter SD, et al. Higher vs lower positive end-expiratory pressure in patients with acute lung injury and acute respiratory distress syndrome: systematic review and meta-analysis. *JAMA : the journal of the American Medical Association*. 2010 Mar 3;303(9):865-73. PubMed PMID: 20197533. Epub 2010/03/04.
8. Briel M. Higher versus lower positive end-expiratory pressure in patients with acute lung injury: a systematic review and individual patient data meta-analysis (study protocol) [Master Thesis]. Hamilton, Canada: McMaster University; 2008.
9. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Statistics in medicine*. 2006 Jan 15;25(1):127-41. PubMed PMID: 16217841.
10. Royston P, Sauerbrei W. A new approach to modelling interactions between treatment and continuous covariates in clinical trials by using fractional polynomials. *Statistics in medicine*. 2004 Aug 30;23(16):2509-25. PubMed PMID: 15287081. Epub 2004/08/03.
11. Royston P, Sauerbrei W. Interactions between treatment and continuous covariates: a step toward individualizing therapy. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2008 Mar 20;26(9):1397-9. PubMed PMID: 18349388.

12. Sauerbrei W, Royston P. A new strategy for meta-analysis of continuous covariates in observational studies. *Statistics in medicine*. 2011 Dec 10;30(28):3341-60. PubMed PMID: 21953493. Epub 2011/09/29.
13. Kasenda B, Sauerbrei W, Royston P, Briel M. Investigation of Continuous Effect Modifiers in a meta-analysis on higher versus lower PEEP in patients requiring mechanical ventilation – protocol of the ICEM study. *Systematic Reviews* (in press). 2014.
14. Brower RG, Lanken PN, MacIntyre N, Matthay MA, Morris A, Ancukiewicz M, et al. Higher versus lower positive end-expiratory pressures in patients with the acute respiratory distress syndrome. *The New England journal of medicine*. 2004 Jul 22;351(4):327-36. PubMed PMID: 15269312. Epub 2004/07/23.
15. Meade MO, Cook DJ, Guyatt GH, Slutsky AS, Arabi YM, Cooper DJ, et al. Ventilation strategy using low tidal volumes, recruitment maneuvers, and high positive end-expiratory pressure for acute lung injury and acute respiratory distress syndrome: a randomized controlled trial. *JAMA : the journal of the American Medical Association*. 2008 Feb 13;299(6):637-45. PubMed PMID: 18270352. Epub 2008/02/14.
16. Mercat A, Richard JC, Vielle B, Jaber S, Osman D, Diehl JL, et al. Positive end-expiratory pressure setting in adults with acute lung injury and acute respiratory distress syndrome: a randomized controlled trial. *JAMA : the journal of the American Medical Association*. 2008 Feb 13;299(6):646-55. PubMed PMID: 18270353. Epub 2008/02/14.
17. Wolkewitz M, Beyersmann J, Gastmeier P, Schumacher M. Modeling the effect of time-dependent exposure on intensive care unit mortality. *Intensive care medicine*. 2009 May;35(5):826-32. PubMed PMID: 19183943. Epub 2009/02/03.
18. Beck DH, Smith GB, Pappachan JV, Millar B. External validation of the SAPS II, APACHE II and APACHE III prognostic models in South England: a multicentre study. *Intensive care medicine*. 2003 Feb;29(2):249-56. PubMed PMID: 12536271. Epub 2003/01/22.
19. Suistomaa M, Niskanen M, Kari A, Hynynen M, Takala J. Customized prediction models based on APACHE II and SAPS II scores in patients with prolonged length of stay in the ICU. *Intensive care medicine*. 2002 Apr;28(4):479-85. PubMed PMID: 11967604. Epub 2002/04/23.
20. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in medicine*. 2011 Feb 20;30(4):377-99. PubMed PMID: 21225900. Epub 2011/01/13.
21. Flegal KM, Carroll MD, Ogden CL, Curtin LR. Prevalence and trends in obesity among US adults, 1999-2008. *JAMA : the journal of the American Medical Association*. 2010 Jan 20;303(3):235-41. PubMed PMID: 20071471.

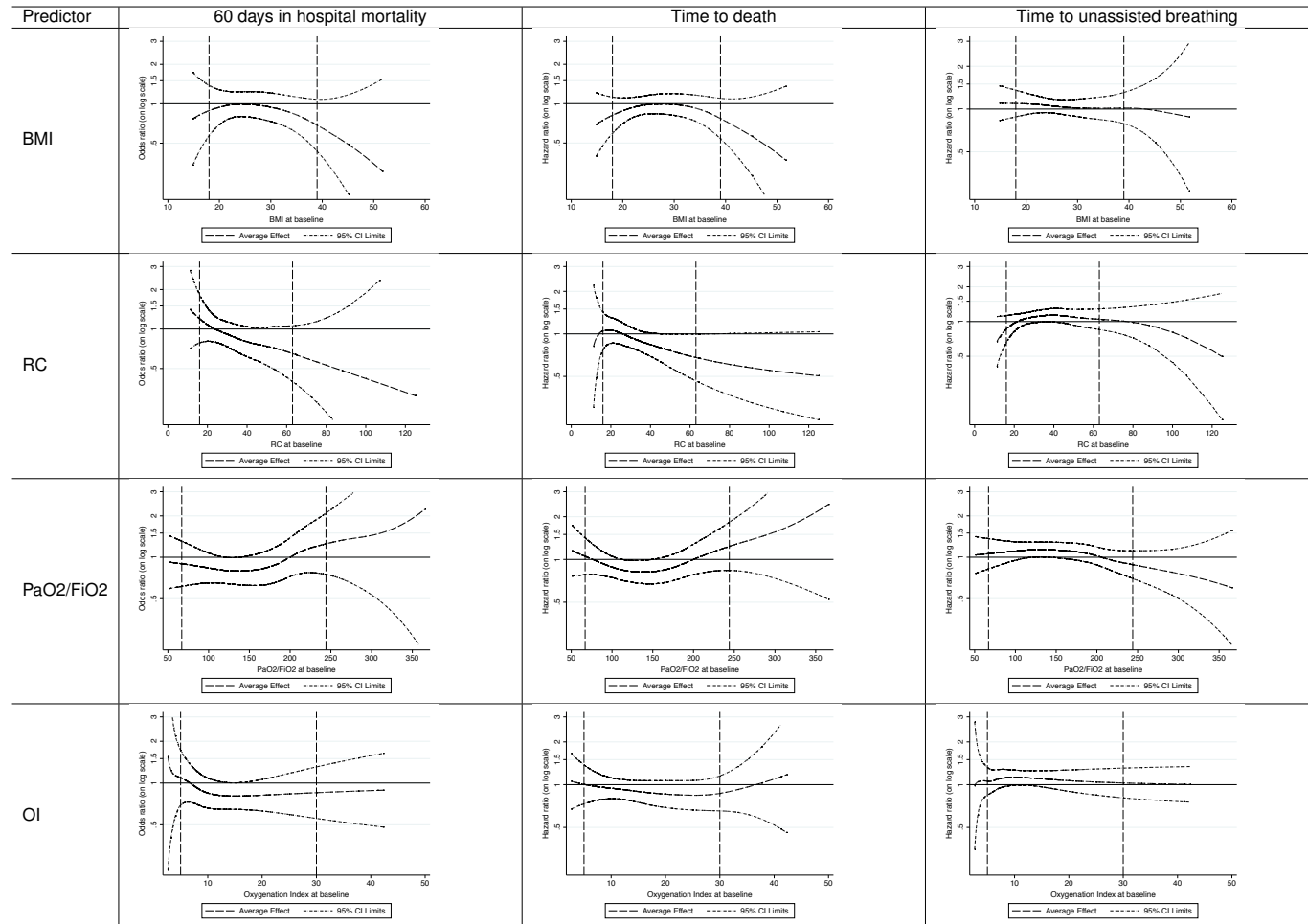
## Tables and Figures

**Table 1:** Selected patient characteristics and endpoints. Values are means (SD) unless specified otherwise. Summary statistics of the characteristics are based on the imputed dataset. Abbreviations: BMI, body mass index; RC, respiratory compliance [tidal volume in ml / inspiratory plateau pressure-PEEP in mmHg]; PEEP, positive end-expiratory pressure; PaO<sub>2</sub>, arterial partial oxygen pressure (mmHg); FiO<sub>2</sub>, fraction of inspired oxygen. \*According to APACHE II and SAPS scores

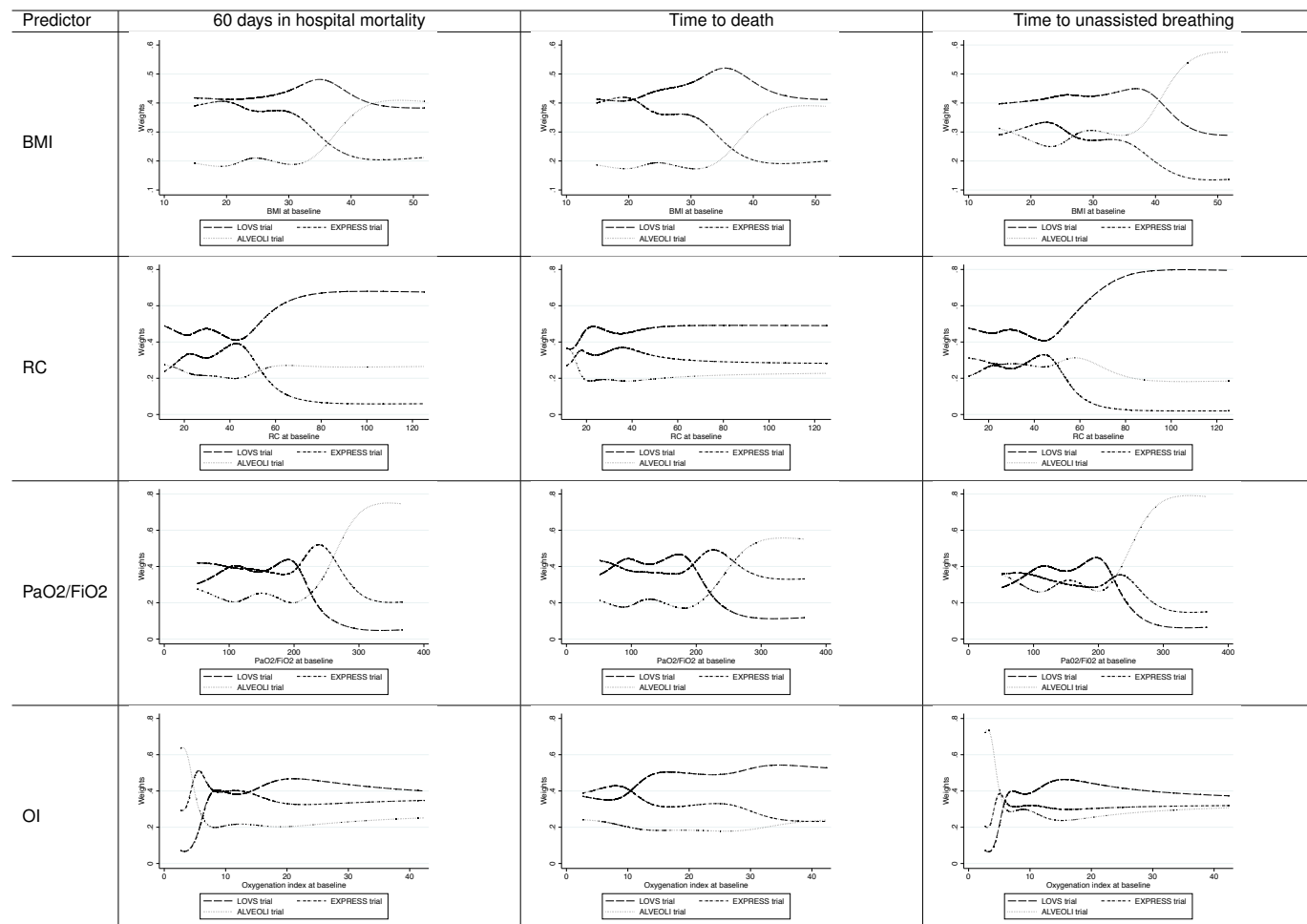
Characteristics	LOVS		EXPRESS		ALVEOLI		ALL	
	Higher PEEP N=475	Lower PEEP N=508	Higher PEEP N=385	Lower PEEP N=382	Higher PEEP N=276	Lower PEEP N=273	Higher PEEP N=1136	Lower PEEP N=1163
Age	54.5 (16.5)	56.9 (16.5)	60.1 (15.7)	59.7 (15.1)	53.7 (17.1)	48.5 (16.9)	56.2 (16.6)	55.9 (16.7)
Women, N (%)	193 (40.6)	201 (39.5)	125 (32.9)	126 (33.2)	119 (43.1)	128 (46.9)	437 (38.5)	455 (39.1)
BMI (kg/m <sup>2</sup> )	27.5 (6.5)	27.2 (6.8)	26.3 (5.8)	26.3 (6.1)	27.8 (6.8)	27.2 (7.0)	27.2 (6.4)	26.9 (6.6)
Missing values, N (%)	72 (15.2)	85 (16.7)	17 (4.4)	19 (5.0)	23 (8.3)	21 (7.7)	112 (9.6)	125 (10.7)
PaO <sub>2</sub> /FiO <sub>2</sub>	145.1 (48.3)	144.7 (49.1)	144 (57.6)	142.7 (56.9)	151 (67.3)	163 (76.2)	146 (56.6)	148.4 (59.5)
Missing values, N (%)	1 (0.2)	2 (0.4)	0 (0)	0 (0)	0 (0)	0 (0)	1 (0.1)	2 (0.2)
Oxygenation index	14.4 (8.3)	14.4 (8.1)	13.1 (7.6)	13.0 (8.2)	13.1 (8.5)	12.1 (8.0)	13.6 (8.1)	13.4 (8.2)
Missing values, N (%)	51 (10.7)	63 (12.4)	80 (20.8)	73 (19.3)	16 (5.8)	18 (6.6)	147 (12.9)	154 (13.2)
Estimated RC (ml/cm H <sub>2</sub> O)	33.4 (16.9)	35.8 (21.6)	33.0 (12.3)	33.0 (11.6)	36.0 (20.7)	31.2 (12.8)	33.9 (16.6)	33.8 (17.1)
Missing values, N (%)	148 (31.2)	170 (21.1)	12 (3.1)	12 (3.1)	34 (12.3)	35 (12.8)	194 (17.1)	217 (18.7)
Probability of death*	52.9 (23.5)	55.7 (23.2)	45.8 (26.6)	44.5 (27.0)	49.0 (21.8)	47.1 (21.1)	49.6 (24.4)	50.0 (24.6)
Missing values, N (%)	0 (0)	0 (0)	0 (0)	0 (0)	3 (1.1)	3 (1.1)	3 (0.3)	3 (0.3)
Severe sepsis, N (%)	214 (45.1)	248 (48.7)	285 (75.0)	268 (70.7)	96 (34.8)	112 (41.0)	595 (52.4)	628 (54.0)
Deaths in hospital, N (%)	162 (34.1)	192 (37.7)	136 (35.8)	149 (39.3)	76 (27.5)	68 (24.9)	374 (32.9)	409 (35.2)
Pneumothorax, N (%)	45 (9.5)	38 (7.5)	26 (6.8)	22 (5.8)	16 (5.8)	15 (5.5)	87 (7.7)	75 (6.4)



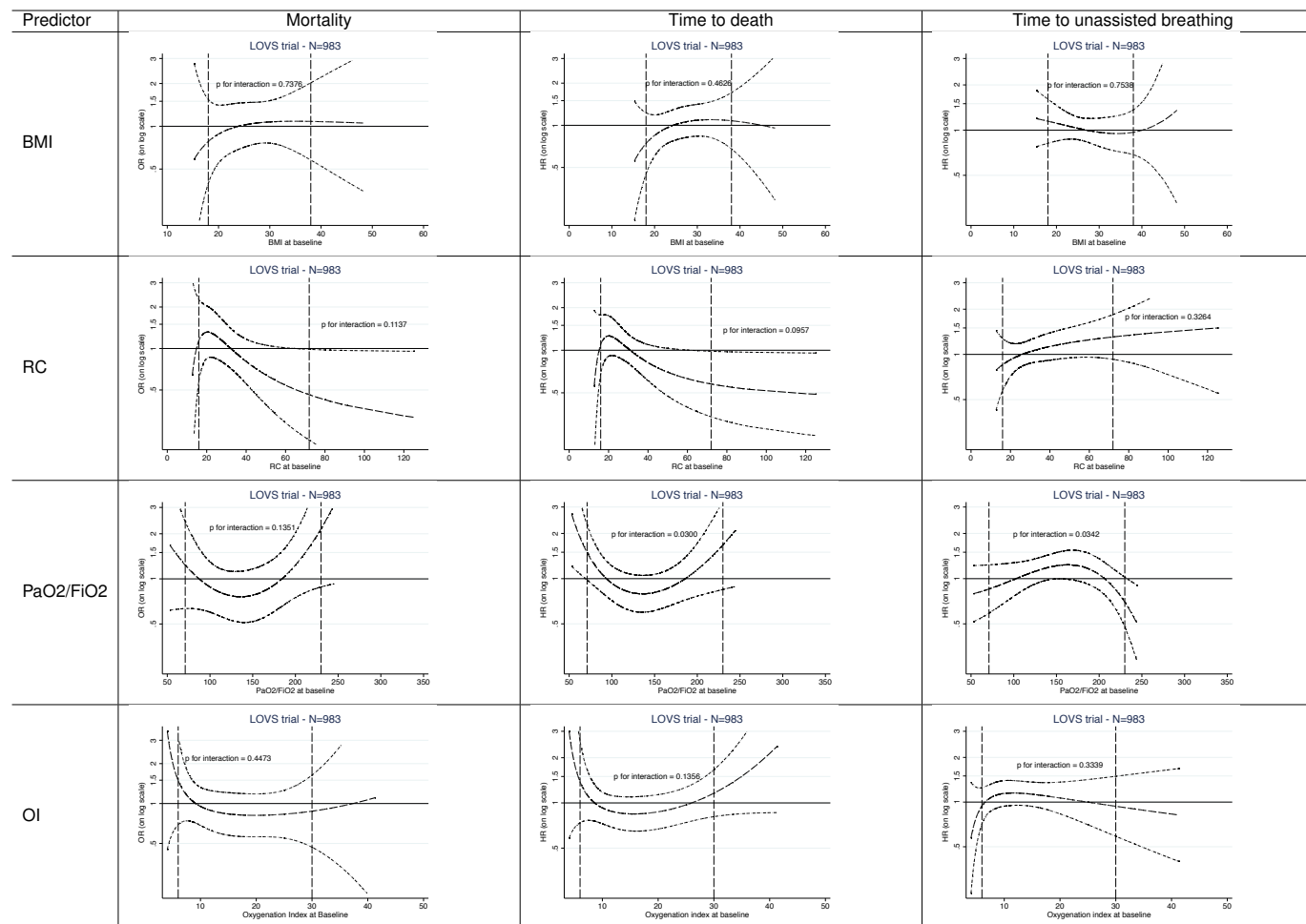
**FIGURE 1:** Trial flow, adapted from Briel et al (7).



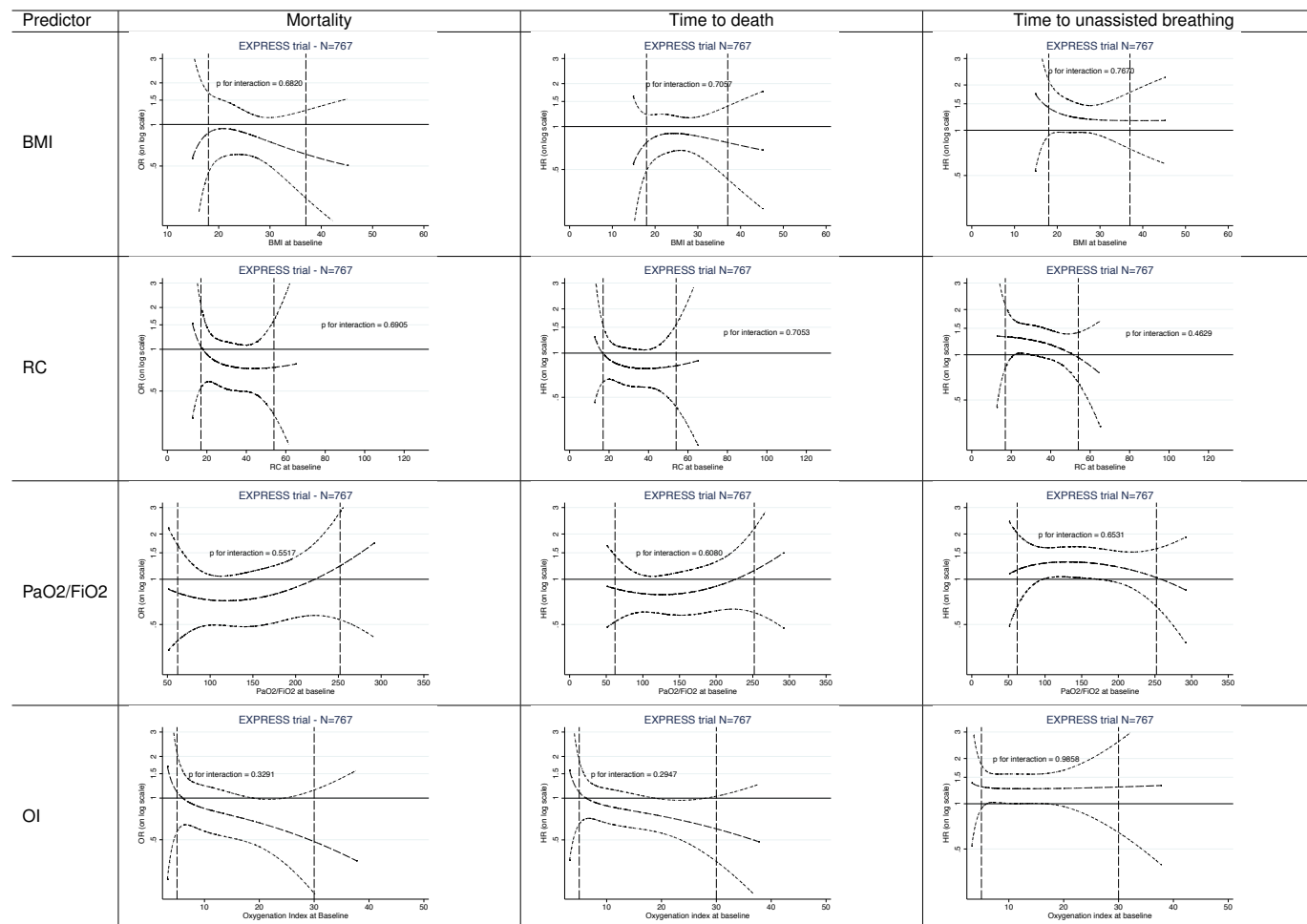
**FIGURE 2:** Summary TEFs based on fixed effects of each predictor-outcome pair. The vertical dashed lines include 95% of the data of the continuous predictors; the horizontal line at the OR or HR of 1 denotes equivalence of treatment effects, thus a TEF parallel to the horizontal line indicates no treatment interaction. For the outcomes 60 days in hospital mortality and time to death values beneath this line indicate that higher PEEP is more effective than lower PEEP. For the outcome time to unassisted breathing, it is the other way round. Abbreviations: CI=confidence interval; BMI=body mass index; HR=hazard ratio; OR=odds ratio; OI=Oxygenation index; PaO<sub>2</sub>/FiO<sub>2</sub>=PaO<sub>2</sub>/FiO<sub>2</sub> ratio; PEEP=Positive endexpiratory pressure; RC=Respiratory compliance; TEF=Treatment effect function



**FIGURE 3:** The respective graphs illustrate the fixed-effects weights for averaging the functions across studies. Weights were derived from the reciprocal of the variances. The shape of these curves is a result of the distribution of events by the respective modifier. Abbreviations: BMI=body mass index; OI=Oxygenation index; PaO<sub>2</sub>/FiO<sub>2</sub>=PaO<sub>2</sub>/FiO<sub>2</sub> ratio; RC=Respiratory compliance; TEF=Treatment effect function

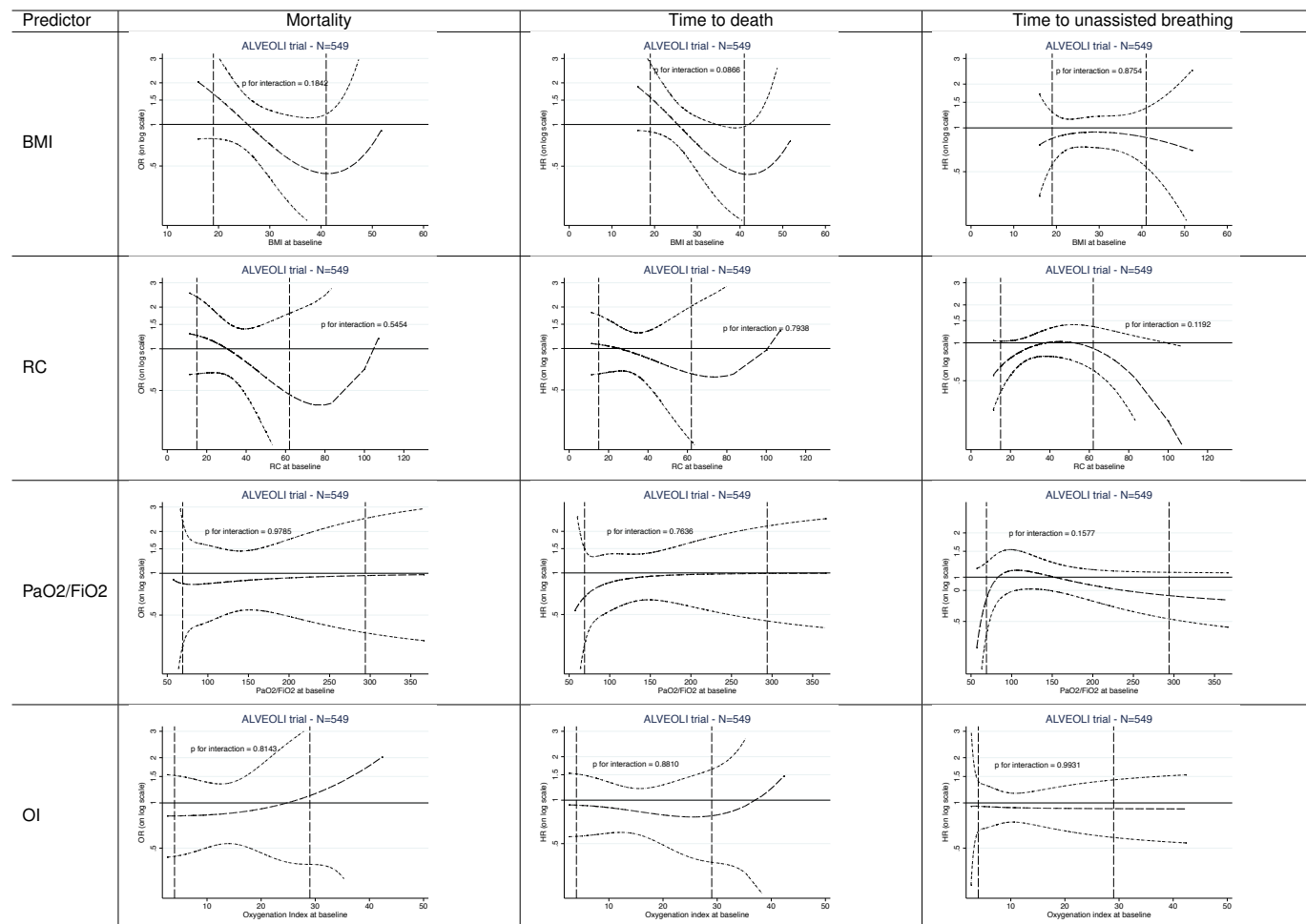


**APPENDIX FIGURE 1:** TEFs of each predictor-outcome pair. The vertical dashed lines include 95% of the data of the continuous predictors; the horizontal line at the OR or HR of 1 denotes equivalence of treatment effects, thus a TEF parallel to the horizontal line indicates no treatment interaction. For the outcomes 60 days in hospital mortality and time to death values beneath this line indicate that higher PEEP is more effective than lower PEEP. For the outcome time to unassisted breathing, it is the other way round. Abbreviations: CI=confidence interval; BMI=body mass index; HR=hazard ratio; OR=odds ratio; OI=Oxygenation index; PaO2/FiO2=PaO2/FiO2 ratio; PEEP=Positive endexpiratory pressure; RC=Respiratory compliance; TEF=Treatment effect function



**APPENDIX FIGURE 2:** TEFs of each predictor-outcome pair. The vertical dashed lines include 95% of the data of the continuous predictors; the horizontal line at the OR or HR of 1 denotes equivalence of treatment effects, thus a TEF parallel to the horizontal line indicates no treatment interaction. For the outcomes 60 days in hospital mortality and time to death values beneath this line indicate that higher PEEP is more effective than lower PEEP. For the outcome time to unassisted breathing, it is the other way round. Abbreviations: CI=confidence interval; BMI=body mass index; HR=hazard ratio; OR=odds ratio; OI=Oxygenation index; PaO2/FiO2=PaO2/FiO2 ratio; PEEP=Positive endexpiratory pressure; RC=Respiratory compliance; TEF=Treatment effect function





**APPENDIX FIGURE 3:** TEFs of each predictor-outcome pair. The vertical dashed lines include 95% of the data of the continuous predictors; the horizontal line at the OR or HR of 1 denotes equivalence of treatment effects, thus a TEF parallel to the horizontal line indicates no treatment interaction. For the outcomes 60 days in hospital mortality and time to death values beneath this line indicate that higher PEEP is more effective than lower PEEP. For the outcome time to unassisted breathing, it is the other way round. Abbreviations: CI=confidence interval; BMI=body mass index; HR=hazard ratio; OR=odds ratio; OI=Oxygenation index; PaO2/FiO2=PaO2/FiO2 ratio; PEEP=Positive end-expiratory pressure; RC=Respiratory compliance; TEF=Treatment effect function

## Further manuscripts during PhD time

---

### Published Original Articles

#### Survival in Overweight Patients with Advanced Pancreatic Carcinoma: A Multicentre Cohort Study

Benjamin Kasenda, Annatina Bass, Dieter Koeberle, Bernhard Pestalozzi, Markus Borner, Richard Herrmann, Lorenz Jost, Andreas Lohri, Viviane Hess

BMC Cancer. 2014 Sep 29;14:728

**BACKGROUND.** Obesity is a risk factor for developing pancreatic cancer. We investigated the impact of obesity on survival in patients diagnosed with locally advanced or metastatic pancreatic cancer.

**PATIENTS AND METHODS.** In a multicentre, retrospective study, we included all patients with advanced or metastatic pancreatic cancer treated at four Swiss hospitals between 1994 and 2004. We categorized patients into four body mass index (BMI) groups (<18.5, 18.5 – 25, ≥ 25 – 29, ≥30 kg/m<sup>2</sup>) and used multivariable Cox regression to investigate the impact of BMI on survival. Missing data were handled using multiple imputations.

**RESULTS.** 483 patients were included. Median age was 66 years (range 59 - 74), 47% were female, 82% had stage IV disease, 72% had an performance status below 2, and 84% were treated with gemcitabine-based first-line chemotherapy. After a median follow-up of 8.5 months, 6 and 12-month survival probabilities of the whole cohort were 67% (95% CI 63% - 71%) and 37% (95% CI 33% - 42%), respectively. Unadjusted 12-month survival rates in each BMI group were: 48% (95% CI 33% - 62%), 42% (95% CI 36% - 48%), 30% (95% CI 22% - 38%), and 11% (95% CI 4% - 24%), respectively. In multivariable analysis, increasing BMI (HR 1.22, 95% CI 1.04 – 1.41, p = 0.012) and CA 19-9 (HR 1.07, 95% CI 1.02 – 1.11, p = 0.003) were significantly associated with worse survival prognosis. Patients with a good clinical performance status (ECOG < 2) had a better prognosis (HR 0.76, 95% CI 0.65 – 0.96, p = 0.019).

**CONCLUSIONS.** Obese patients diagnosed with advanced pancreatic cancers have a worse prognosis compared to non-obese patients. BMI should be considered for risk stratification in future clinical trials.

## Prevalence, Characteristics, and Publication of Discontinued Randomized Trials

**Benjamin Kasenda**, Erik von Elm, John, Anette Blümle, Yuki Tomonaga, Ramon Saccilotto, Alain Amstutz, Theresa Bengough, Joerg J. Meerpohl, Mihaela Stegert, Kari A. O. Tikkinen, Ignacio Neumann, Alonso Carrasco-Labra, Markus Faulhaber, Sohail M. Mulla, Dominik Mertz, Elie A. Akl, Dirk Bassler, Jason W Busse, Ignacio Ferreira-González, Francois Lamontagne, Alain Nordmann, Viktoria Gloy, Heike Raatz, Lorenzo Moja, Rachel Rosenthal, Shanil Ebrahim, Stefan Schandelmaier, Xin Sun, Per O. Vandvik, Bradley C. Johnston, Martin A. Walter, Bernard Burnand, Matthias Schwenkglenks, Lars G. Hemkens, Heiner C. Bucher, Gordon H. Guyatt, and Matthias Briel

JAMA 2014; 311(10): 1045-1051

**IMPORTANCE.** The discontinuation of randomized controlled trials (RCTs) raises ethical concerns and often wastes scarce research resources. The epidemiology of discontinued RCTs, however, remains unclear.

**OBJECTIVES.** To determine the prevalence, characteristics, and publication history of discontinued RCTs, and to investigate factors associated with RCT discontinuation due to poor recruitment and with non-publication.

**DESIGN AND SETTING.** Retrospective cohort of RCTs based on archived protocols approved by six research ethics committees (RECs) in Switzerland, Germany, and Canada between 2000 and 2003. We recorded trial characteristics and planned recruitment from included protocols. Last follow-up of RCTs was April 27th 2013.

**MAIN OUTCOMES.** Completion status, reported reasons for discontinuation, and publication status of RCTs as determined based on filed correspondence with RECs, literature searches, and investigator surveys.

**RESULTS.** After a median follow-up of 11.6 years (range, 8.8–12.6 years), 253 of 1017 included RCTs were discontinued (24.9%, 95% confidence interval [CI], 22.3%-27.6%). Only 96 of 253 discontinuations (37.9%, 95% CI, 32.0%-44.3%) were reported to RECs. Most frequent reason for discontinuation was poor recruitment (101/1017; 9.9%, 95% CI, 8.2%-12.0%). In multivariable analysis, industry- versus investigator-sponsorship (7.8% versus 27.2%, odds ratio [OR] 0.23, 95% CI, 0.14 – 0.40;  $p < 0.001$ ) and a larger planned sample size [increments of 100] (OR 0.96, 95% CI 0.92 – 1.00,  $p = 0.044$ ) were associated with lower rates of discontinuation due to poor recruitment. Discontinued trials were more likely to remain unpublished than completed trials (55.1% versus 33.6%; OR 3.22, 95% CI, 2.32–4.49;  $p < 0.001$ ).

**CONCLUSION AND RELEVANCE.** In this sample of trials based on RCT protocols from six RECs, discontinuation was common, with poor recruitment being the most frequently reported reason. Greater efforts are needed to ensure the reporting of trial discontinuation to RECs and the publication of results of discontinued trials.

## **<sup>18</sup>F-FDG PET Is an Independent Outcome Predictor in Primary Central Nervous System Lymphoma**

Benjamin Kasenda, Vanessa Haug, Elisabeth Schorb, Kristina Fritsch, Jürgen Finke, Michael Mix, Claudia Hader, Wolfgang A. Weber, Gerald Illerhaus, and Philipp T. Meyer

Journal of Nuclear Medicine 2013 Feb; 54(2): 184-91

**BACKGROUND:** Primary central nervous system (CNS) lymphoma is an aggressive non-Hodgkin lymphoma with poor prognosis. We evaluated pretreatment <sup>18</sup>F-FDG PET as a prognostic marker in primary CNS lymphoma.

**PATIENTS AND METHODS:** Forty-two immunocompetent patients with newly diagnosed primary CNS lymphoma who underwent pretreatment <sup>18</sup>F-FDG PET were retrospectively analysed. Baseline status and response to treatment were evaluated by MR imaging. Tumour maximum standardized uptake values were assessed by volume-of-interest analyses using an automatic isocontour definition. A 10-step semiquantitative visual rating system (metabolic imaging lymphoma aggressiveness scale, or MILAS) was used to assess primary CNS lymphoma metabolism as a marker of clinical aggressiveness. Logistic regression, log-rank testing, and multivariable Cox regression were used to investigate the association between <sup>18</sup>F-FDG uptake and tumor response and survival.

**RESULTS:** Mean maximum standardized uptake value correlated linearly with MILAS. The distribution of patients according to MILAS (0-9) was 0%, 28.6%, 23.8%, 21.4%, 11.9%, 4.8%, 7.1%, 0%, 0%, and 2.4%. There was no correlation between MILAS and response to treatment. Respective 2- and 5-y survival rates were 52% and 32% for progression-free survival (PFS) and 64% and 50% for overall survival (OS). A cutoff at MILAS 3 was a good separator for PFS (median: 54.7 mo [ $\leq 3$ ], 3.8 mo [ $> 3$ ],  $P = 0.0272$ ) and OS (median: not reached [ $\leq 3$ ], 13.8 mo [ $> 3$ ],  $P = 0.131$ ). In multivariable analyses, increasing MILAS was significantly associated with shorter PFS (hazard ratio, 1.49,  $P = 0.006$ ) and OS (hazard ratio, 1.43,  $P = 0.018$ ).

**CONCLUSION:** Increased pretreatment <sup>18</sup>F-FDG uptake may offer new opportunities for baseline risk evaluation in untreated primary CNS lymphoma.

## Prognosis of patients with primary central nervous system lymphoma after high-dose chemotherapy followed by autologous stem cell transplantation.

Elisabeth Schorb, Benjamin Kasenda, Johannes Atta, Stephan Kaun, Anke Morgner, Georg Hess, Thomas Elter, Nikolas von Bubnoff, Martin Dreyling, Mark Ringhoffer, Stefan W. Krause, Günter Derigs, Beate Klimm, D. Niemann, Kristina Fritsch, Jürgen Finke, and Gerald Illerhaus

Haematologica. 2013 May; 98(5): 765-70

**BACKGROUND:** High-dose chemotherapy followed by autologous stem cell transplantation has been shown to be feasible and highly effective in newly diagnosed primary central nervous system lymphoma. In this retrospective multicentre study we investigated prognosis and baseline risk factors in patients with primary central nervous system lymphoma who underwent this treatment approach.

**PATIENTS AND METHODS:** We retrospectively analysed 105 immunocompetent patients with primary central nervous system lymphoma who underwent high-dose chemotherapy followed by autologous stem cell transplantation with or without whole brain radiotherapy as first line consolidation treated at 12 German centres between 1997 and 2011. We estimated survival rates and investigated the impact of age, performance status, serum lactate dehydrogenase level, and deep brain involvement on overall and progression-free survival. Patients were additionally categorized into three prognostic groups according to the Memorial Sloan Kettering Cancer Centre prognostic model.

**RESULTS:** After a median follow-up of 47 months, median progression free survival and overall survival was reached after 85 and 121 months; 2 and 5-years overall survival rates were 82% and 79%, respectively. The Memorial Sloan Kettering Cancer Centre prognostic model did not predict survival. Only age revealed some evidence of prognostic relevance. Overall response rate was 95%; of those patients with progressive disease before high-dose chemotherapy, 7/20 achieved ongoing complete remission after therapy without whole brain radiation therapy. Transplantation-associated mortality was 2.8%.

**CONCLUSIONS:** High-dose chemotherapy followed by autologous stem cell transplantation is a highly effective and safe treatment modality for selected primary central nervous system lymphoma patients. Superiority compared to standard chemotherapy still warrants further investigation.

## Prognosis after high-dose chemotherapy followed by autologous stem-cell transplantation as first-line treatment in primary CNS lymphoma - a long-term follow-up study.

Benjamin Kasenda, Elisabeth Schorb, Kristina Fritsch, Jürgen Finke, and Gerald Illerhaus

Annals of Oncology 2012 Oct; 23(10): 2670-5

**BACKGROUND:** High-dose chemotherapy followed by autologous stem-cell transplantation (HCT-ASCT) is a promising approach in eligible patients with primary central nervous system lymphoma (PCNSL).

**PATIENTS AND METHODS:** We report long-term data of patients who were treated according to HCT-ASCT containing protocols. Patients and methods We analyzed survival and relapse rates in 43 (<67 years) immunocompetent patients with newly diagnosed PCNSL being treated according to two different high-dose methotrexate-based protocols followed by high-dose carmustine/thiotepa (BCNU/TT) plus ASCT ( $\pm$ whole brain irradiation). Analysis was conducted for all patients (intention-to-treat) and those patients who actually received HCT-ASCT (per-protocol).

**RESULTS:** Thirty-four patients achieved complete remission, of those 12 relapsed (35%), while 6 of them relapsed 5 years after diagnosis. After a median follow-up of 120 months, median overall survival (OS) was reached after 104 months. Two- and 5-year OS was 81% and 70% and 2- and 5-year event-free survival (EFS) was 81% and 67%, respectively. In per-protocol analysis (N = 34), 5-year OS and EFS was 82% and 79%, respectively. HCT-ASCT associated related mortality was not observed.

**CONCLUSIONS:** Sequential high-dose MTX containing chemotherapy followed by high-dose carmustine/thiotepa plus ASCT ( $\pm$ whole brain irradiation) is safe and leads to high survival rates in eligible patients with newly diagnosed PCNSL.

## The prognostic value of serum methotrexate area under curve in elderly primary CNS lymphoma patients

Benjamin Kasenda, Marcel Rehberg, Petra Thürmann, Melanie Franzem, Hendrik Veelken, Kristina Fritsch, Elisabeth Schorb, Jürgen Finke, Dirk Lebiecz, and Gerald Illerhaus

Annals of Hematology 2012 Aug; 91(8): 1257-64

**BACKGROUND:** Studies on pharmacokinetics and pharmacodynamics of high-dose methotrexate chemotherapy (HD-MTX) in elderly primary central nervous system lymphoma (PCNSL) patients are rare. MTX exposure time has recently been proposed as an outcome determining factor in PCNSL.

**PATIENTS AND METHODS:** We investigated 49 immunocompetent PCNSL patients (female N=30, male N=19, median age 73 years) who were treated according to HD-MTX-based protocols. A two-compartment pharmacokinetic model was used to describe the MTX clearance. Response to treatment was assessed by MRI. We used multivariable models to investigate the association between MTX exposure and tumor response as well as survival.

**RESULTS:** Dose normalized MTX peak serum levels [C (max),  $\mu\text{mol/L g}$ ] and dose normalized area under the curve [AUC(dn),  $\mu\text{mol h/L g}$ ] were higher in females than in males, respectively [59.4 (f) vs. 48.1 (m),  $P < 0.001$ ; 373.2 (f) vs. 271.9 (m),  $P = 0.008$ ]. Increasing AUC was inversely correlated with tumor response. AUC values above 2,126 h  $\mu\text{mol/L}$  were independently associated with shorter overall and progression-free survival [hazard ratio (HR), 4.56, 95 % CI 1.74-11.94; HR 2.87, 95 % CI 1.18-7.00].

**CONCLUSIONS:** Exceedingly high MTX AUC levels can have a negative impact on progression-free and overall survivals in elderly PCNSL patients.

## Original Articles accepted for publication

### Completion and publication rates of randomized controlled trials in surgery - an empirical study

Rachel Rosenthal, **Benjamin Kasenda**, Salome Dell-Kuster, Erik von Elm, John You, Anette Blümle, Yuki Tomonaga, Ramon Saccilotto, Alain Amstutz, Theresa Bengough, Jörg Meerpohl, Mihaela Stegert, Kari A. O. Tikkinen, Ignacio Neumann, Alonso Carrasco-Labra, Markus Faulhaber, Sohail Mulla, Dominik Mertz, Elie A. Akl, Dirk Bassler, Jason Busse, Ignacio Ferreira-González, Francois Lamontagne, Alain Nordmann, Viktoria Gloy, Heike Raatz, Lorenzo Moja, Shanil Ebrahim, Stefan Schandelmaier, Xin Sun, Per O. Vandvik, Bradley C. Johnston, Martin A. Walter, Bernard Burnand, Matthias Schwenkglens, Lars Hemkens, Heiner C. Bucher, Gordon H. Guyatt, and Matthias Briel

Accepted for publication in *Annals of Surgery* (03/2014)

**OBJECTIVE:** To investigate the prevalence of discontinuation and non-publication of surgical versus medical randomized controlled trials (RCTs) and to explore risk factors for discontinuation and non-publication of surgical RCTs.

**SUMMARY BACKGROUND DATA:** Trial discontinuation has significant scientific, ethical, and economic implications. To date, the prevalence of discontinuation of surgical RCTs is unknown.

**METHODS:** All RCT protocols approved 2000-2003 by six ethics committees in Canada, Germany and Switzerland were screened. Baseline characteristics were collected and, if published, full reports retrieved. Risk factors for early discontinuation for slow recruitment and non-publication were explored using multivariable logistic regression analyses.

**RESULTS:** In total, 863 RCT protocols involving adult patients were identified, 127 in surgery (15%) and 736 in medicine (85%). Surgical trials were discontinued for any reason more often than medical trials (43% *versus* 27%, risk difference 16% (95% confidence interval [CI] 5%, 26%);  $p=0.001$ ) and more often discontinued for slow recruitment (18% *versus* 11%, risk difference 8% (95% CI 0.1%, 16%);  $p=0.020$ ). The percentage of trials not published as full journal article was similar in surgical and medical trials (44% *versus* 40%, risk difference 4% (95% CI -5%, 14%);  $p=0.373$ ). Discontinuation of surgical trials was a strong risk factor for non-publication (odds ratio 4.18, 95% CI 1.45, 12.06;  $p=0.008$ ).

**CONCLUSIONS:** Discontinuation and non-publication rates were substantial in surgical RCTs and trial discontinuation was strongly associated with non-publication. These findings need to be taken into account when interpreting surgical literature. Surgical trialists should consider feasibility studies before embarking on full-scale trials.



## Original Articles currently under review for publication

### First-Line Treatment and Outcome of Elderly Patients with Primary Central Nervous System Lymphoma (PCNSL) – A Systematic Review and Individual Patient Data Meta-Analysis

**Benjamin Kasenda**, Andrés JM Ferreri, Emerenziana Marturano, Deborah Forst, Jacoline Bromberg, Hervé Ghesquieres, Celine Ferlay, Jean Yves Blay, Khe Hoang Xuan, Yasushi Okoshi, Shigeru Chiba, Kristina Fritsch, Antonio Omuro, Brian Patrick O'Neill, Osnat Bairey, Stefan Schandelmaier, Viktoria Gloy, Neera Bhatnagar, Stefan Haug, Susanne Rahner, Tracy T Batchelor, Gerald Illerhaus, and Matthias Briel

Submitted to Annals of Oncology [11/2014]

**PURPOSE.** To investigate prognosis and effects of first-line therapy in elderly primary central nervous system lymphoma (PCNSL) patients.

**PATIENTS AND METHODS.** A systematic review of studies about first-line therapy in immunocompetent patients  $\geq 60$  years with PCNSL until 2013 and a meta-analysis of individual patient data from eligible studies and international collaborators were performed.

**RESULTS.** We identified 17 eligible studies; from 12 studies we obtained individual data of 346 patients which were pooled with data of 395 additional patients (N=741). Median age and Karnofsky Performance Score (KPS) was 68 years (range: 60 - 90) and 60% (10% - 100%), respectively. KPS  $\geq 70\%$  was the strongest prognostic factor for mortality (HR 0.52, 95% CI 0.42 - 0.64). After a median follow-up of 44 months, 2-year survival was 42% (95% CI, 38 - 46). 276 patients received whole brain radiotherapy (WBRT) (median 36 Gy, range 28.5 - 70); 51% received a dose  $> 36$  Gy. High-dose methotrexate (HD-MTX)-based therapy was associated with improved survival (HR 0.70, 95% CI 0.53 - 0.93). There was no difference between HD-MTX plus oral chemotherapy only and more aggressive HD-MTX based therapies (HR 1.45, 95% CI 0.95 - 2.24). WBRT seemed to improve survival, but was associated with an increase for neurotoxicity (Odds ratio 5.56, 95% CI 2.47 - 12.45).

**CONCLUSIONS.** Elderly PCNSL patients benefit from HD-MTX. More aggressive HD-MTX protocols do not seem to improve outcome. WBRT was associated with improved survival, but neurotoxicity remains a concern. Randomized trials for elderly PCNSL patients are warranted.

## Published Reviews

### Lung cancer screening - an overview about chances and risks

Benjamin Kasenda, Heike Raatz, and Heiner Bucher

Therapeutische Umschau 2013 Apr; 70(4): 237-43

**BACKGROUND:** Lung cancer is a leading cause of death worldwide. Patients are usually diagnosed at an advanced stage and have a very poor prognosis. In Switzerland, lung cancer is the most frequent cause of cancer death in men and the second most frequent cause of cancer death in women. Programmes to prevent individuals from initiating to smoke and to support smokers to quit are the most effective lung cancer prevention strategy. Whether routine screening for lung cancer in smokers is effective to reduce lung cancer related morbidity and mortality remains questionable.

**METHODS:** We summarize the evidence of five recent randomised controlled trials on routine screening for lung cancer in smokers.

**RESULTS:** One study found no benefit of periodic conventional chest X-rays as compared to usual care without regular imaging for reducing lung cancer death. In four other trials, low-dose computer tomography (LDCT) was compared to conventional chest X-rays and to usual care. Only the largest trial, the US based National Lung Cancer Screening Trial (NLST), demonstrated a statistically significant reduction of lung cancer mortality of LDCT compared to conventional chest X-rays whereas three European trials could not prove any benefit.

**CONCLUSIONS:** The results of the NLST need to be interpreted with care due to limited generalizability to European settings. LDCT screening had an unacceptable high rate of false positive findings resulting in an enormous use of resources for diagnostic work-up. Whether LDCT screening is associated with an acceptable incremental cost-effectiveness ratio still warrants further investigation.

## Meta-analyses: what they can and cannot do.

Alain Nordmann, Benjamin Kasenda, and Matthias Briel

Swiss Medical Weekly. 2012 Mar 9; 142:w13518

Meta-analyses overcome the limitation of small sample sizes or rare outcomes by pooling results from a number of individual studies to generate a single best estimate. As long as a meta-analysis is not limited by poor quality of included trials, unexplainable heterogeneity and/or reporting bias of individual trials, meta-analyses can be instrumental in reliably demonstrating benefit or harm of an intervention when results of individual randomised controlled trials are conflicting or inconclusive. Therefore meta-analyses should be conducted as part of a systematic review, i.e., a systematic approach to answer a focused clinical question. Important features of a systematic review are a comprehensive, reproducible search for primary studies, selection of studies using clear and transparent eligibility criteria, standardised critical appraisal of studies for quality, and investigation of heterogeneity among included studies. Cumulative meta-analysis may prevent delays in the introduction of effective treatments and may allow for early detection of harmful effects of interventions. As opposed to meta-analysis based on aggregate study data, individual patient data meta-analyses offer the advantage to use standardised criteria across trials and reliably investigate subgroup effects of interventions. Network meta-analysis allows the integration of data from direct and indirect comparisons in order to compare multiple treatments in a comprehensive analysis and determine the best treatment among several options. We conclude that meta-analysis has become a popular, versatile, and powerful tool. If rigorously conducted as part of a systematic review, it is essential for evidence-based decision making in clinical practice as well as on the health policy level.

## Discussion and Future Steps

---

*Instead, I want to rescue the scientific importance of valid pathophysiologic subgroups from being forgotten or destroyed by excessive vehemence in suggestions that all subgroups are evil. The essence of tragedy has been described as the destructive collision of two sets of protagonists, both of whom are correct. The statisticians are right in denouncing subgroups that are formed post hoc from exercises in pure data dredging. The clinicians are also right, however, in insisting that a subgroup is respectable and worthwhile when established a priori from pathophysiologic principles.*

Alvan R Feinstein, Journal of Clinical Epidemiology 1998 (1)

Individualized or personalized medicine has become a buzzword in the academic as well as public debate surrounding health care. Based on a recent systematic review, the following definition for personalized medicine has been proposed: “Personalized medicine seeks to improve stratification and timing of health care by utilizing biological information and biomarkers on the level of molecular disease pathways, genetics, proteomics as well as metabolomics.” (2) Following this definition, the concept of personalized medicine is not necessarily about persons - it’s about subgroups and the more refined nosology of modern medicine which is based on much more profound knowledge on the pathological processes. Therefore, a more technical, but probably more adequate term would be stratified medicine (3), because personalization or individualization of medicine is much more an aspect of including patients’ values and preferences into clinical decision making which goes beyond simple application of guidelines and measuring biomarkers. However, the less technical term *personalized medicine* is probably more appealing, because it better transports the message of a new medicine evoking hopes for patients and physicians.

Biotechnologies are rapidly emerging and especially in oncology, immense efforts are made to identify molecular targets in tumours to develop a *mechanism-based therapy* – treatments are tailored to target driving pathomechanisms of a malignancy (4) with impressive improvements in some entities e.g. chronic myeloid leukaemia and aggressive B-cell lymphoma (5, 6). Other examples highlighting the impact of targeted agents comprise the treatment of breast cancer depending on human epidermal growth factor - 2 (HER-2) expression status, genetic expression levels (7-9) or metastatic adenocarcinoma of the lung depending on epidermal growth factor receptor (EGFR) driver mutation status (10). Therefore, common categorization of many tumours, which are traditionally based on microscopic morphology and site of origin, are no longer sufficient, because modern treatment strategies now require more information.

Because targeted agents are developed to address specific disease-driving factors, clinicians and patients hope for increasing treatment efficacy while reducing toxicity. Many targeted agents, however, do not work in all patients simply because the knowledge about the disease mechanism to target is not fully understood – current models about the mechanism

only partly explain the clinical activity. This can have great impact on patient care, because these targeted agents may not only be ineffective, but also harmful. For example, EGFR antibodies (cetuximab and panitumumab) are now used in addition to palliative chemotherapy in metastatic adenocarcinoma of the colon. These antibodies increase tumour response and progression free survival rates in patients harbouring a K-RAS wild type adenocarcinoma of the colon (11-13). Only recently new evidence indicated that the K-RAS status, and other mutations in the RAS gene family have an obverse clinical relevance. Those patients with RAS wild type benefit from addition of anti-EGFR antibodies, patients whose tumours harbour mutations in the RAS gene family are harmed – they die earlier (14)! Therefore, although mechanisms of disease are increasingly better understood, one has to keep in mind that the promise of efficacy without side effects is hard to hold given that knowledge about the known disease specific mechanisms only partly explain activity of the drug. New clinical trial designs using e.g. an adaptive randomization in early trial stages may allow an earlier selection of patients who are more likely to benefit from the targeted agents (15).

### Subgroup effects – trust and techniques

Sun et al proposed an approach to account for the uncertainty of inference from results of subgroup analyses. Based on previous work (16) 11 credibility criteria have been proposed to help judging the likelihood of a subgroup effect to be true in RCT reports or meta-analyses of RCTs (17). It is unlikely that a subgroup claim will meet either all or none of these proposed criteria—it is rather much more likely, that a subgroup claim will meet some but not all the criteria. This means that the greater the extent to which the criteria are met, the more may clinicians or health policy makers believe in the observed subgroup effect, which in turn influences their clinical or political decisions. One of the central criteria is the pre-specification of a subgroup analysis with an underlying hypothesis, which is emphasized by various experts (18-20). However, our results challenge this central key criterion, because the quality of subgroup planning/reporting in RCTs is still very limited. We therefore conclude that unless a reliable source such as a trial protocol is available, readers of RCT reports should view statements about subgroup effects with great scepticism, even if the authors state that the subgroup analysis had been pre-specified.

Categorizing an inherently continuous variable raises several critical issues in statistical analyses. Conclusions based on such analyses can be wrong with subsequent impact on treatment recommendations and decisions. I applied a new statistical technique, the multivariable fractional polynomial interaction approach, to investigate interaction in an individual patient data meta-analysis. Results suggest that ventilated patients with PaO<sub>2</sub>/FiO<sub>2</sub> between 100 and 150mmHg benefit most from higher PEEP ventilation strategies with respect to 60 days in-hospital mortality, time-to-death, and time-to-unassisted breathing. Also patients with a respiratory compliance above 40 ml/cm H<sub>2</sub>O may derive a benefit from higher PEPP levels. If IPD are available, the MFPI is a straightforward method to investigate interactions between continuous predictors and outcomes by using all information available in the dataset. For all these analyses, I received great support from Willi Sauerbrei and Patrick Royston, the founders of the MFPI approach. It was therefore the

first time that this kind of analyses was conducted with real patient data in the framework of an individual patient data meta-analysis of RCTs. As outlined before, pre-specification of a subgroup is one central criterion to judge the credibility of a subgroup effect (17), but why? It is simply to account for a false discovery rate of significant effects that in fact may be wrong (type 1 error). This “fear” of false positive subgroup effects is of most concern especially to analysts of medical data. However, Sauerbrei and Royston recently pointed out that the *“excessive focus on controlling the false discovery rate has detracted attention away from the problem of an excessive type 2 error rate or ‘false non-discovery’ rate. Overlooking a clinically relevant treatment–marker interaction is a serious error and may lead to inappropriate treatment of patients. To reduce the chance of false-negative findings, we must develop, evaluate, and apply good statistical methods to detect interactions. The power to detect interactions with continuous covariates can be increased by analysis on the original scale. We therefore contend that MFPI, which does this, is such a method.”* (21) Still, also MFPI analysis should be planned and testing hypotheses should always be preceded by careful reasoning about the assumptions made.

### Future projects

To investigate the appraisal of subgroup effects on another level, it would be interesting to know to what extent results from subgroup analyses are included in clinical guidelines of different medical societies. The GRADE Working Group has established a systematic approach to appraise the level of evidence, which is especially of interest to authors of clinical guidelines (22). However, results from subgroup analysis are not explicitly included in this approach. As outlined above, subgroup analyses in RCTs and meta-analysis are common and can have great impact on clinical decision-making. In oncology, clinical guidelines often include results from subgroup analyses to provide guidance in clinical decision-making (23–25) and it is very likely that results from subgroup analyses will increasingly be considered in these guidelines given the fast development of identifying new disease classifications and treatment approaches for cancer patients (10, 14, 26–28). In this emerging field, it would be interesting to investigate how authors of these guidelines appraise the confidence in subgroup results and whether there is an association between the yet available proposed credibility criteria (17) and recommendations made by the guideline authors. A research project investigating this question would comprise the following work packages: 1. Select an international oncology society that regularly publishes updated clinical guidelines (e.g. European Society of Medical Oncology), 2. Select a sample of the most current guidelines on each tumour entity, 3. Identify treatment recommendations that are based on subgroup analysis and overall effects from RCTs, 4. Investigate to which extent the 11 credibility criteria are met, 5. Appraise the association between fulfilment of the criteria and strength of recommendations based on subgroup analysis. Results from such a project would provide first evidence on which criteria results from subgroup analysis are implemented in guidelines for cancer treatment and likely reveal inappropriate confidence in results from subgroup analysis.

However, by just pointing out problems in appraising subgroup effects, things are unlikely to improve. Although the proposed 11 credibility criteria are backed by statistical principles,

some examples, and thorough discussions among clinical epidemiologists over the last decades, the meta-epidemiological evidence for the association between the grade of fulfilment of these criteria and the confidence in the truth of a subgroup effect, has yet not been provided. Furthermore, also the single relative weights of each criterion used to appraise a subgroup effect are far from being quantified. One possible criterion for high confidence in reported subgroup effect is the confirmation in an independent dataset – a very basic but utmost important principle in clinical research, but not often at hand especially for subgroup analyses. Thus it would be desirable to provide clinicians and researchers with a reliable and valid instrument for assessing the credibility of subgroup effects so that they can optimally tailor care to individual patients. To develop such an instruments, the following work packages would be needed: 1. Systematically review previous methods for the interpretation of subgroup analyses in RCTs, 2. Identify items that inform the credibility of a subgroup effect, 3. Develop an instrument to measure the credibility of subgroup effects, 4. Define threshold scores for the new instrument to guide treatment decisions based on putative subgroup effects, and 5. Examine sensibility of the instrument by external experts and conduct an empirical assessment test of its reliability and validity. Such a study would address the question that clinicians and health policy makers face every day - to what extent should research results be applied across a broad population? Current frameworks for making this assessment are inconsistent, and no tool with demonstrated reliability or validity is yet available to help judge subgroup credibility. If a valid instrument could be developed, it may also be implemented in approaches such as GRADE.

Subgroup analyses are mostly conducted on the primary or secondary efficacy endpoints – safety is rarely considered as the primary endpoint. Composite endpoints such as cardiovascular morbidity (time from randomization to myocardial infarction, ischemic stroke, or death) or progression free survival (time from randomization until death, tumour progression, or relapse) are common composite endpoints used in cardiovascular or oncology trials (29). Reasons for choosing composite endpoints are the higher rate of clinical events compared to a single event such as death, thus the number of required events is reached within a shorter time interval with less patients to prove a hypothesis of clinical efficacy (29). Further issues associated with such composite endpoints are discussed elsewhere (30). Safety outcomes are usually compared between randomized groups using absolute frequencies and percentages; however, little is known on how often safety/tolerability outcomes are analysed within subgroups. Because the clinical benefit of cancer therapies in palliative settings are often small, benefit and harms have to be carefully balanced. Therefore it is not only important to identify those subgroups that benefit most regarding a specific efficacy outcome, but also those who are likely to have less severe side effects that interfere with patients' all day life. One problem that arises to answer this question is that safety endpoints are very heterogeneous and usually categorized according to body system as proposed in the well-established common toxicity criteria (31, 32). Despite the heterogeneity of safety outcome types, the common toxicity criteria allow to rank the severity of outcomes on a scale from 0 to 5 (0 denotes no side effect and 5 denotes death due to the side effect). If a side effect clearly interferes negatively with patients' all day life, it is at least ranked as grade three. To improve patient care, it would be desirable to

have better estimates for possible severe side effects. Because the number of events of specific side effects is likely too low in common RCTs it would be necessary to define well-accepted composite safety outcomes similar to efficacy outcomes. A project investigating this issue would comprise the following work packages: 1. Systematically review previous methods for appraising safety outcomes in oncology trials, 2. Select a sample of meta-analysis in oncology trials including patients with advanced cancer to identify the most frequent safety endpoints and their incidences, 3. Group these safety endpoints according to transient and long-term, 4. Define clinical useful composite safety endpoints, 5. Define potential vulnerable subgroups of patients, 6. Collect individual patient data if available from included studies of the selected meta-analysis to conduct pre specified subgroup analysis for the proposed safety composite endpoints of interest.

As mentioned previously, based on our updated literature search and personal communication, there is one ongoing RCT (NTC00431158) likely being eligible for inclusion in our data set of critically ill patients requiring mechanical ventilation. Therefore, in a next step we will contact the primary investigator of that trial and invite him to collaborate by providing individual patient data to be pooled with our set. This might provide us with more power to show a stable interaction effect between BMI and ventilation strategy using the MFPI approach. Given that the primary investigator of the still ongoing RCT would agree to share individual patient data after completion of the trial, such a project of incorporating this additional RCT would include the following working packages: 1. Defining the variables of interest, 2. Harmonizing the definitions and units of the variables with our dataset, 3. Checking for missing values and impute missing values, 3. Combining the datasets, and 4. Re-analysing the dataset with the already pre-defined hypothesis for interaction.

### Closing remarks

Subgroup analyses and hypotheses are of great importance within the continuous scientific process of falsification and verification. Therefore, if subgroup analyses are planned, they should be conceptualized *a priori*, limited in their number and justified like any other hypothesis tested in a trial. Ideally, the sample size of the trial should allow for sufficient power to test for interaction; however, this is un-realistic for most RCTs. It is of course not wrong to conduct subgroup analyses based on post-hoc hypothesis or just with an exploratory or hypothesis-generating intent, but readers should be made aware of limitations and pitfalls - guidelines how to report subgroup analyses have been proposed (19). Besides appropriate statistical techniques, investigators reporting subgroup results should always provide a detailed description about their motivation for their subgroup analyses and also call the reader's attention to external evidence supporting or contradicting their subgroup findings if available.



## References

1. Feinstein AR. The problem of cogent subgroups: a clinicostatistical tragedy. *Journal of clinical epidemiology*. 1998 Apr;51(4):297-9. PubMed PMID: 9539884. Epub 1998/04/16. eng.
2. Schleidgen S, Klingler C, Bertram T, Rogowski WH, Marckmann G. What is personalized medicine: sharpening a vague term based on a systematic literature review. *BMC medical ethics*. 2013;14:55. PubMed PMID: 24359531. Pubmed Central PMCID: 3878093.
3. Hingorani AD, Windt DA, Riley RD, Abrams K, Moons KG, Steyerberg EW, et al. Prognosis research strategy (PROGRESS) 4: stratified medicine research. *Bmj*. 2013;346:e5793. PubMed PMID: 23386361. Pubmed Central PMCID: 3565686.
4. Wistuba, II, Gelovani JG, Jacoby JJ, Davis SE, Herbst RS. Methodological and practical challenges for personalized cancer therapies. *Nature reviews Clinical oncology*. 2011 Mar;8(3):135-41. PubMed PMID: 21364686.
5. O'Brien SG, Guilhot F, Larson RA, Gathmann I, Baccarani M, Cervantes F, et al. Imatinib compared with interferon and low-dose cytarabine for newly diagnosed chronic-phase chronic myeloid leukemia. *The New England journal of medicine*. 2003 Mar 13;348(11):994-1004. PubMed PMID: 12637609.
6. Coiffier B, Lepage E, Briere J, Herbrecht R, Tilly H, Bouabdallah R, et al. CHOP chemotherapy plus rituximab compared with CHOP alone in elderly patients with diffuse large-B-cell lymphoma. *The New England journal of medicine*. 2002 Jan 24;346(4):235-42. PubMed PMID: 11807147.
7. Senkus E, Kyriakides S, Penault-Llorca F, Poortmans P, Thompson A, Zackrisson S, et al. Primary breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO*. 2013 Oct;24 Suppl 6:vi7-23. PubMed PMID: 23970019.
8. Slamon D, Eiermann W, Robert N, Pienkowski T, Martin M, Press M, et al. Adjuvant trastuzumab in HER2-positive breast cancer. *The New England journal of medicine*. 2011 Oct 6;365(14):1273-83. PubMed PMID: 21991949. Pubmed Central PMCID: 3268553.
9. Paik S, Tang G, Shak S, Kim C, Baker J, Kim W, et al. Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2006 Aug 10;24(23):3726-34. PubMed PMID: 16720680.
10. Lee CK, Brown C, Gralla RJ, Hirsh V, Thongprasert S, Tsai CM, et al. Impact of EGFR inhibitor in non-small cell lung cancer on progression-free and overall survival: a meta-analysis. *Journal of the National Cancer Institute*. 2013 May 1;105(9):595-605. PubMed PMID: 23594426.
11. Karapetis CS, Khambata-Ford S, Jonker DJ, O'Callaghan CJ, Tu D, Tebbutt NC, et al. K-ras mutations and benefit from cetuximab in advanced colorectal cancer. *The New England journal of medicine*. 2008 Oct 23;359(17):1757-65. PubMed PMID: 18946061.
12. Bokemeyer C, Van Cutsem E, Rougier P, Ciardiello F, Heeger S, Schlichting M, et al. Addition of cetuximab to chemotherapy as first-line treatment for KRAS wild-type metastatic colorectal cancer: pooled analysis of the CRYSTAL and OPUS randomised clinical trials. *European journal of cancer*. 2012 Jul;48(10):1466-75. PubMed PMID: 22446022.
13. Douillard JY, Siena S, Cassidy J, Tabernero J, Burkes R, Barugel M, et al. Randomized, phase III trial of panitumumab with infusional fluorouracil, leucovorin, and oxaliplatin (FOLFOX4) versus FOLFOX4 alone as first-line treatment in patients with previously untreated metastatic colorectal cancer: the PRIME study. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2010 Nov 1;28(31):4697-705. PubMed PMID: 20921465.
14. Douillard JY, Oliner KS, Siena S, Tabernero J, Burkes R, Barugel M, et al. Panitumumab-FOLFOX4 treatment and RAS mutations in colorectal cancer. *The New England journal of medicine*. 2013 Sep 12;369(11):1023-34. PubMed PMID: 24024839.

15. Zhou X, Liu S, Kim ES, Herbst RS, Lee JJ. Bayesian adaptive design for targeted therapy development in lung cancer--a step toward personalized medicine. *Clinical trials*. 2008;5(3):181-93. PubMed PMID: 18559407.
16. Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. *Annals of internal medicine*. 1992 Jan 1;116(1):78-84. PubMed PMID: 1530753. Epub 1992/01/01.
17. Sun X, Briel M, Walter SD, Guyatt GH. Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. *Bmj*. 2010;340:c117. PubMed PMID: 20354011. Epub 2010/04/01.
18. Rothwell PM. Treating individuals 2. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. *Lancet*. 2005 Jan 8-14;365(9454):176-86. PubMed PMID: 15639301. Epub 2005/01/11. eng.
19. Wang R, Lagakos SW, Ware JH, Hunter DJ, Drazen JM. Statistics in medicine--reporting of subgroup analyses in clinical trials. *The New England journal of medicine*. 2007 Nov 22;357(21):2189-94. PubMed PMID: 18032770. Epub 2007/11/23. eng.
20. Yusuf S, Wittes J, Probstfield J, Tyroler HA. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. *JAMA : the journal of the American Medical Association*. 1991 Jul 3;266(1):93-8. PubMed PMID: 2046134. Epub 1991/07/03. eng.
21. Royston P, Sauerbrei W. Interaction of treatment with a continuous variable: simulation study of significance level for several methods of analysis. *Statistics in medicine*. 2013 Sep 30;32(22):3788-803. PubMed PMID: 23580422.
22. Guyatt GH, Oxman AD, Kunz R, Vist GE, Falck-Ytter Y, Schunemann HJ, et al. What is "quality of evidence" and why is it important to clinicians? *Bmj*. 2008 May 3;336(7651):995-8. PubMed PMID: 18456631. Pubmed Central PMCID: 2364804.
23. Dummer R, Hauschild A, Guggenheim M, Keilholz U, Pentheroudakis G, Group EGW. Cutaneous melanoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO*. 2012 Oct;23 Suppl 7:vii86-91. PubMed PMID: 22997461.
24. Group EESNW. Gastrointestinal stromal tumors: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO*. 2012 Oct;23 Suppl 7:vii49-55. PubMed PMID: 22997454.
25. Peters S, Adjei AA, Gridelli C, Reck M, Kerr K, Felip E, et al. Metastatic non-small-cell lung cancer (NSCLC): ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO*. 2012 Oct;23 Suppl 7:vii56-64. PubMed PMID: 22997455.
26. Shrager J, Tenenbaum JM. Rapid learning for precision oncology. *Nature reviews Clinical oncology*. 2014 Jan 21. PubMed PMID: 24445514.
27. Dancey J. Targeted therapies and clinical trials in ovarian cancer. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO*. 2013 Dec;24 Suppl 10:x59-x63. PubMed PMID: 24265407.
28. Shaw AT, Kim DW, Nakagawa K, Seto T, Crino L, Ahn MJ, et al. Crizotinib versus chemotherapy in advanced ALK-positive lung cancer. *The New England journal of medicine*. 2013 Jun 20;368(25):2385-94. PubMed PMID: 23724913.
29. Freemantle N, Calvert M, Wood J, Eastaugh J, Griffin C. Composite outcomes in randomized trials: greater precision but with greater uncertainty? *JAMA : the journal of the American Medical Association*. 2003 May 21;289(19):2554-9. PubMed PMID: 12759327.
30. Montori VM, Permyer-Miralda G, Ferreira-Gonzalez I, Busse JW, Pacheco-Huergo V, Bryant D, et al. Validity of composite end points in clinical trials. *Bmj*. 2005 Mar 12;330(7491):594-6. PubMed PMID: 15761002. Pubmed Central PMCID: 554039.

31. Trotti A, Colevas AD, Setser A, Rusch V, Jaques D, Budach V, et al. CTCAE v3.0: development of a comprehensive grading system for the adverse effects of cancer treatment. *Semin Radiat Oncol*. 2003 Jul;13(3):176-81. PubMed PMID: 12903007.
32. Trotti A, Colevas AD, Setser A, Basch E. Patient-reported outcomes and the evolution of adverse event reporting in oncology. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2007 Nov 10;25(32):5121-7. PubMed PMID: 17991931.

# Curriculum Vitae

**Dr. med. Benjamin Kasenda**  
**born 28th September 1980 in Datteln**  
**Germany**

**Work.** Medical Oncology, Royal Marsden Hospital, London, UK

[benjamin.kasenda@rmh.nhs.uk](mailto:benjamin.kasenda@rmh.nhs.uk)

**Private.** 32 Albert Palace Mansions, Lurline Gardens, SW11 4DG London, UK

[benjamin.kasenda@gmail.com](mailto:benjamin.kasenda@gmail.com),  
 +44-(0)79 30 488 046

## Profession

07/2014-present	Clinical Research Fellow, Medical Oncology, Royal Marsden Hospital, London, UK
02/2013-06/2014	Clinical Fellow in Oncology, Department of Medical Oncology, University Hospital Basel, Switzerland
11/2010-01/2013	Junior Researcher, Basel Institute for Clinical Epidemiology and Biostatistics, Switzerland
05/2010-10/2010	Resident, Emergency Medicine (Dep. of Hematology and Oncology, Freiburg University Medical Center, Germany)
10/2009-04/2010	Resident, general internal medicine and nephrology (Dep. of Nephrology and General Internal Medicine, Freiburg University Medical Center, Germany)
04/2009-09/2009	Resident, hematopoietic stem cell transplantation (Dep. of Hematology and Oncology, Freiburg University Medical Center, Germany)
10/2008-03/2009	Resident, medical oncology (Dep. of Hematology and Oncology, Freiburg University Medical Center, Germany)

## Medical Education

11/2014	PhD in Epidemiology, University of Basel
06/2008	MD thesis at the Institute of Immunology at Witten/Herdecke University (Magna cum Laude)
05/2008	License to practice medicine
04/2002	Enrollment at Witten/Herdecke University, Germany

## Teaching Activity

11/2010 – 01/2013	Teaching for medical undergraduates in Clinical Epidemiology, University of Basel
04/2003- 08/2006	Assistant at the chair for Medical Ethics and Communication Witten/Herdecke University
04/2004- 06/2006	Tutor in Problem Based Learning for medical students
04/2003-11/2005	Teaching Wu Style Tai Chi Quan in the framework of the Studium fundamentale at Witten/Herdecke University

## Scholarships

06/2014	SAKK / Janssen Fellowship
10/2004- 05/2008	Foundationer of Studienstiftung des deutschen Volkes
since 05/2008	Alumnus of Studienstiftung des deutschen Volkes

## Memberships

Since 03/2013	European Association of Medical Oncology (ESMO)
Since 12/2012	German Association for Hematology/Oncology (DGHO)
06/2011-06/2014	Professional body of the German physicians of Internal Medicine

## Languages

German	First language
English	Very good in word and writing
Chinese	Good in word and basic writing
French	Good in word good writing