

INVITED COMMENTARY

STATISTICAL PRINCIPLES FOR PROSPECTIVE STUDY
PROTOCOLS:

DESIGN, ANALYSIS, AND REPORTING

Robin Christensen, BSc, MSc, PhD^{1,2}Henning Langberg, PT, MSc, DMSc³

ABSTRACT

In the design of scientific studies it is essential to decide on which scientific questions one aims to answer, just as it is important to decide on the correct statistical methods to use to answer these questions. The correct use of statistical methods is crucial in all aspects of research to quantify relationships in data. Despite an increased focus on statistical content and complexity of biomedical research these topics remain difficult for most researchers. Statistical methods enable researchers to condense large spreadsheets with data into means, proportions, and difference between means, risk differences, and other quantities that convey information. One of the goals in biomedical research is to develop parsimonious models - meaning as simple as possible. This approach is valid if the subsequent research report (the article) is written independent of whether the results are "statistically significant" or not. In the present paper we outline the considerations and suggestions on how to build a trial protocol, with an emphasis on having a rigorous protocol stage, always leading to a full article manuscript, independent of statistical findings. We conclude that authors, who find (rigorous) protocol writing too troublesome, will realize that they have already written the first half of the final paper if they follow these recommendations; authors simply need to change the protocols future tense into past tense. Thus, the aim of this clinical commentary is to describe and explain the statistical principles for trial protocols in terms of design, analysis, and reporting of findings.

Key words: analysis, research design, statistics

CORRESPONDING AUTHOR

Robin Christensen, MSc, PhD;

Senior Biostatistician

Associate Professor of Statistics in Medicine

Head of Musculoskeletal Statistics Unit,

The Parker Institute, Department of

Rheumatology, Copenhagen University

Hospital, Frederiksberg, Copenhagen F,

Denmark

E-mail: Robin.Christensen@Frh.Regionh.DK;

Fax: + 45 3816 4159.

¹ Musculoskeletal Statistics Unit, The Parker Institute, Department of Rheumatology, Copenhagen University Hospital, Frederiksberg, Copenhagen F, Denmark.

² Institute of Sports Science and Clinical Biomechanics, University of Southern Denmark, Odense, Denmark.

³ CopenRehab, Institute of Social Medicine, Department of Public Health and Centre for Healthy Ageing, Faculty of Health Sciences, University of Copenhagen, Denmark.

INTRODUCTION

Generalized linear models, sequential analysis, time series, survival analysis, design of experiments, residuals and diagnostics, likelihood inference and statistical approximation – are all topics that might cause confusion and distract the consumer of a research paper, rather than help to clarify the objective of a biomedical research project. Despite the fact that the statistical content and complexity of biomedical research has increased steadily over recent decades,¹ with reports of clinical trials containing a wealth of data comparing treatments,² there is still a need to educate researchers in order to allow for “parsimonious statistical thinking”. Researchers, without a solid knowledge of clinical epidemiology and/or biostatistics, are often increasing the complexity of their model in their eagerness to explain “everything”, assuming that a statistician will “show up later” and “sort the whole thing out”. One of the goals in biomedical research is to develop parsimonious models to allow the testing of a hypothesis to be as simple as possible. The idea of parsimony, is here used synonymous with “Occam’s razor”, and states that when *“everything else is equal, simple models are to be preferred over complex models”*.

Applied statistics should not focus too much on distributions or probabilities, but rather focus on the ‘*a priori*’ approach to “telling the good story”; that is, focus – *a priori* - on how to present the project when the data collection has been finalized, and analyzed. When designing a research study it is important to decide on the statistical methods that will apply *before* starting the data collection. Researchers, who cannot disclose the anticipated outline of their paper *a priori*, are most likely introducing bias into the biomedical literature.² The end user of biomedical research, might infer erroneous conclusions (after synthesis) from the totality of the available evidence, as a consequence of the complete results of all conducted studies on a question of interest not being available in the public domain, thereby introducing publication and/or selective outcome reporting bias.³ Biostatistics is the use of numbers to quantify relationships in data, whether empirical or causal, and thereby answer the scientific questions raised in the original protocol. This approach is valid if the subsequent research article is written independent of whether the results are “statistically significant”

or not. It is very unfortunate, that the process of running the main analyses requires only little time, whereas preparing a manuscript requires considerable effort, and thus frequently leads to situations where the scientists ponder (*post hoc*) whether it is worth manuscript preparation. They weigh the perceived importance and priority of the question, the statistical significance of any (quickly obtained) results, and other evidence circulating at the time.⁴

Statistical methods should enable the researcher to reduce a large spreadsheet with collected variables (including outcome measures and design variables) into means, proportions, and difference between means, risk differences, and other quantities that convey information. In principle this kind of statistics are called descriptive statistics, enabling the reader to do most of the subsequent statistical tests by hand while reading the paper⁵ a phenomenon often referred to as transparent reporting of statistical data. The use of explicit numerical information will be enhanced if authors remain focused on the original ‘Statistical Analysis Plan’ (SAP) as outlined in the pre-specified protocol of the study.

The authors objective for this manuscript is to provide suggestions and considerations on what to prepare before initiating the study (collecting data), with an emphasis on having a rigorous protocol stage that leads to a full article manuscript, independent of statistical findings.

CONCEPTUALIZING COHORT STUDIES AS IF THEY WERE RCTS

Valid evidence on the benefits and risks of healthcare interventions is essential in decision-making. Randomized controlled trials (RCTs) are considered the ultimate method for providing evidence on efficacy. Frequently, however, the RCTs are criticized for focusing on highly selected populations and outcomes. Therefore cohort studies can be thought of as natural experiments in which outcomes are measured in real world rather than in experimental settings. Whether the statistical design is based on observational or randomized data, the reporting of both types of studies is often of insufficient quality, and poor reporting hampers the assessment of the strengths and weaknesses of a study and the generalizability of its results. Because of this, a group of scientists and editors devel-

oped the CONSORT (Consolidated Standards of Reporting Trials) statement to improve the quality of reporting of RCTs,⁶ and the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) statement with recommendations on how to improve the quality when reporting from observational studies.⁷ The STROBE statement covers cohort, cross-sectional, and case-control designs. Often the ideas and intentions of cohort studies are similar to RCTs in that they compare outcomes in groups that did and did not receive an intervention (Exposed vs. Unexposed, respectively). The main difference is that allocation of individuals is not by chance (random) in the observational studies. Rigorous RCTs eliminate selection bias, balancing both known and unknown prognostic factors, in the assignment of treatments.⁶ If done inadequately, treatment comparisons may be prejudiced, whether consciously or not, by selection of participants of a particular kind to receive a particular treatment.

Unlike RCTs, cohort studies are always vulnerable to selection bias. In cohort studies, factors that determined whether a person received the intervention could result in the groups differing in factors related to the outcome, either because people were preferentially selected to receive one treatment or because of choices that they made. These baseline differences in prognosis could “confound the assessment of the effect” of the intervention. Confounding means “confusion of effects”;⁷ as a consequence, authors could naively consider a cohort study comparing individuals exposed to those individuals without exposure as a quasi-randomized trial when building the SAP. This should then be followed by a clear description about which (potential) confounding factors they would adjust for in the final and fully adjusted model. Information on the distribution of potential confounders in the intervention and comparison groups is usually—and preferably - provided in the first table of a manuscript reporting from a cohort study.

Confounding is a problem only if some baseline characteristics are unevenly distributed between the intervention and comparison groups. When reporting such a study, where you compare groups at baseline (Exposed vs. Unexposed or Intervention vs. Placebo), it is tempting to report a p-value, despite the fact that the test probably either has a low statistical power, or

are impossible to interpret as it is the case for RCTs.⁶ In assessing cohort studies, it is important to identify potential confounders and to examine their distribution in the exposed and control group. Although unknown confounders are always difficult to deal with in cohort studies, a systematic approach can be used to identify known and potential confounders. Typical features could be demographic (e.g., age and sex), medical (e.g., concomitant disease/conditions), previous or current exposure to drugs, and social and behavioral factors (e.g., habits, exercise and diet). When comparing different potential confounding factors between groups at baseline, as previously suggested by Mamdani et al, the authors recommend alternative to traditional significance testing, authors use standardized differences or effect sizes to examine ‘between group differences’ in patient characteristics.⁸ The most important thing in regard to detecting bias due to confounding (assessed at baseline), is that it cannot be judged solely by statistical significance; usually the impact of a potential confounder is judged pragmatically on whether the adjustment for the “confounding variable” changes the estimate of association. The authors of this manuscript recommend that authors of cohort studies to apply a “three-stage reporting framework”: Model #1 report the crude analysis (without any adjustments for confounders); Model #2 report the semi-adjusted analysis (including all the *a priori* defined covariates that are likely to be confounding factors); Model #3 Fully adjusted model (including all the protocolized covariates as in model #2 plus those of potential interest according to baseline table discrepancies across groups).

METHODS SECTIONS IN PROTOCOL

In the following sections the authors will elaborate on important aspects of designing and reporting a two-group comparison whether it is a randomized trial, or a cohort study. For those with a particular interest in cohort studies we will encourage them to scrutinize the highly informative ‘*Reader’s guide to critical appraisal of cohort studies*’ published in 2005 in British Medical Journal.⁹⁻¹⁰ Likewise, we remind authors of observational studies to consult the STROBE statement on how to report cohort, cross-sectional, and case-control studies in biomedical journals.⁷ As a consequence of trying to exclude the technicalities involved in cohort studies, we will in

the latter part focus on what to consider when preparing a report for a controlled trial. Two documents are highly endorsed: The CONSORT statement (used by IJSPT),⁶ as well as the international guideline on statistical methodology (E9) to the collection of 'International Conference on Harmonisation' (ICH E9) on statistical principles for clinical trials.¹¹

Trial protocol

Once a "clinical question" has been raised, the first step in the conception of a trial is to develop a trial protocol. Protocol development has evolved over the last decades. Writing a protocol for a trial is both an intellectual, rigorous, and creative task, which is usually done by researchers and experts in various areas of research including medical, scientific, statistical, ethical, administrative fields, and maybe even regulatory bodies. The protocol serves as the foundation for study conduct and reporting. Full knowledge of a trial protocol allows an appropriate ethical assessment before trial inception, and the proper critical appraisal of the results after trial completion. The SPIRIT initiative (Standard Protocol Items: Recommendations for Interventional Trials) are currently preparing a statement that will most likely include a final 33-item checklist that will apply in the future when researchers design and register their trials. This SPIRIT checklist will support the CONSORT statement, and explicitly ask for details about important features of the pre-specified protocol.

It is recommended that the protocol include at least the following domains: Administrative information; Introduction – including background, rationale, and objective(s); Methods (#1): Participants, interventions, and outcomes; Methods (#2): Assignment of interventions – including sequence generation, concealed allocation, etc.; Methods (#3): Sample size, data collection process, management, and statistical analysis; Methods (#4): Data monitoring; Results: Propose a preliminary outline of the study report; Ethics and dissemination; and finally appendices to all of the above. Independent of these proposed protocol items, we strongly encourage trialists to follow the CONSORT checklist when designing the study.

Before launching a trial, it is important that the researcher considers whether the hypotheses are stated in advance and planned to be evaluated to

confirm the hypotheses (i.e., a confirmatory trial), or whether the researchers want to explore different clear and precise objectives (i.e., exploratory trial). Unlike confirmatory trials, exploratory studies may not always lead to simple tests of predefined hypotheses. With this in mind it seems that exploratory studies, as they are more flexible, may be the "better strategic choice"; however, be aware that an exploratory study can never confirm whether e.g. a physical therapy intervention is effective or not. Decision makers would likely argue, that a given intervention that shows promising results in exploratory studies would need at least one (phase-3-like) confirmatory 'landmark study'.

Trial design

The design of the trial should be described, for example, as parallel group, cluster randomized, crossover, factorial, superiority, equivalence or non-inferiority design, or some other combination of these designs.⁶ The most common RCT design is the parallel group trial, in which participants are randomized to one of two (or more) interventions, with each arm being allocated a different treatment. The assumptions for these trials are less complex than for most other trial designs.

In a cross-over design, participants are randomly allocated to different sequences of treatments, and thereby act as his or her own control for treatment comparisons. In the simplest 2×2 cross-over design each participant receives each of two treatments in a randomized order in two successive periods, usually separated by a washout period. Cross-over designs have a number of caveats that can invalidate their results. The major concern is "carryover effects" where the first treatment given is able to influence the "response" in the subsequent treatment. The statistical consequence might be that the effect of unequal carryover will be to bias direct treatment comparisons.

In a factorial design, participants are assigned to more than one treatment-comparison group, enabling two or more treatments being evaluated simultaneously. The simplest example is the 2×2 factorial design where participants are randomly allocated to one of four possible combinations of two treatments (e.g. A and B). This allows the statistical model to statistically

conclude from A alone; B alone; both A & B; neither A nor B. Usually the statistical focus in studies with factorial design will be on examining the interaction between A and B.

Participants

For each of the following paragraph headings, that are all directly associated with the protocol the authors remind researchers to report what they anticipate will characterize the participants rather than waiting for the final data. This includes explicit eligibility criteria (i.e., both inclusion and exclusion criteria) for participants and the settings where the data will be collected. RCTs address an issue relevant to a particular population or group with the condition of interest. Participant eligibility criteria may relate to demographics, clinical diagnosis, and comorbid conditions. A clear description of the trial participants and setting in which they will be studied is needed to allow future readers to assess the external validity (generalizability) of the trial and determine its applicability to their own setting.

Interventions and blinding

Detailed information on the interventions intended for each group including the essential features of the experimental and comparison interventions (e.g., control group) should be described. Authors should report details about the interventions, e.g., dose, route of administration, duration of administration, surgical procedure, or manufacturer of inserted device. When describing the interventions it is very important to report whether or not participants care givers, and those assessing the outcomes will be blinded to group assignment.¹² Blinding refers to the practice of keeping the trial participants, care providers, data collectors, and sometimes those analyzing the data, unaware of which intervention is being administered to which participant, to reduce the risk of bias. Authors should avoid using terms such as “single” or “double” blind as such terms are not precise enough.

Outcomes

It is important to clearly define what will be the primary outcome for the trial, as well as what are the secondary outcomes. The primary outcome (‘primary variable’, ‘target variable’, ‘primary endpoint’¹³)

should be the variable capable of providing the most clinically relevant and convincing evidence directly related to the primary objective of the trial and is usually the one variable used for the sample size calculation. Most trials have several outcomes, some of which are deemed more important than others.³ Such rankings are typically reported as primary and secondary outcomes. Ideally there should only be one primary outcome. Authors should explicitly state the primary outcome for the trial and when it will be assessed (e.g., the time frame over which it is measured).

Secondary outcomes are either supportive measurements related to the primary objective or measurements of effects related to the secondary objectives. Their definition in the protocol is also important, as well as an explanation of their relative importance and roles in interpretation of trial results. Generally, the number of secondary outcome variables should be limited and should be related to the limited number of questions attempted answered in the trial.

Sample size

The number of participants in a trial should always be large enough to provide a reliable answer to the research questions addressed. The number of participants randomized to each intervention group is an essential element of the results of a trial. This number defines the sample size, and readers of the published article can use it to assess whether all randomized participants were included in the study and subsequent data analysis (referred to as the “Intention To Treat” (ITT) population).¹⁴

Investigators calculate sample sizes before the start of their trial and adequately describe what went into the calculation in details in their protocol as well as their published report. In these *a priori* calculations, determining the effect size to detect—e.g., difference between means ($MD = M_1 - M_C$), or risk difference calculated from the proportions who respond ($RD = p_1 - p_C$)—reflects inherently subjective clinical judgments.¹⁵ The term treatment effect or effect size generally means – for efficacy outcomes – the net benefit, of applying intervention ‘I’ compared to intervention ‘C’. Typically the type of outcome data used for sample size estimation is either ‘binary data’ (where the participant can have a response yes/no),

or it can be estimated from 'continuous data' (where the variable is typically guestimated from an expected mean value and a corresponding standard deviation). The authors need to decide *a priori* what kind of clinical net benefit (difference between groups) they would expect from the intervention, with focus on clinical relevance.

When determining the appropriate sample size, the following items should be considered and specified: (#1) a primary outcome (i.e., the "name" and whether it is binary or continuous by nature); (#2) the test statistic, (#3) the null hypothesis, and the alternative hypothesis (i.e., the reason for the study); (#4) the probability of erroneously rejecting the null hypothesis (the type I error, i.e., the statistical p-value); and (#5) the probability or erroneously failing to reject the null hypothesis (the type II error, i.e., 1 - the type II error is referred to as the statistical power of the study). When determining the sample size for a trial it is also important to consider the protocol approach to dealing with treatment withdrawals and protocol violations.

In an excellent tutorial paper, published in the *Lancet*, Schulz and Grimes argue that the subjective judgments needed from the authors (i.e., content experts) to be able to estimate the sample size - or maybe the statistical power if the number of participants is already determined - is necessary for the trial to be trusted.¹⁵ Realizing that these judgments greatly affect sample size calculations, Schulz and Grimes question the branding of trials as unethical on the basis of an imprecise sample size calculation process. They claim, that, some shift of emphasis from a fixation on sample size to a focus on methodological quality would yield more trials with less bias;¹⁶ unbiased trials with imprecise results trump no results at all.¹⁵

Randomization

Investigators should make sure to describe how participants will be randomized and allocated to the different interventions. It is important to conceal the allocation sequence from those assigning participants to the intervention groups. Allocation concealment prevents investigators from influencing which participants are assigned to a given intervention group (i.e., selection bias). Evidence shows that reports of trials

reporting inadequate allocation concealment are associated with exaggerated treatment effects. Authors should clearly describe in the protocol the method for assigning participants to interventions. Examples of approaches used to ensure adequate concealment include: centralized (e.g., allocation by a central office) or pharmacy-controlled randomization; sequentially numbered identical containers that are administered serially to participants; on-site computer system combined with allocations kept in a locked, unreadable computer file that investigators can access only after the characteristics of an enrolled participant are entered; and sequentially numbered, opaque sealed envelopes.⁶

Statistical methods

Medical research is carried out on selected individuals, although the selection criteria are not always clear. As indicated above, each of the mentioned paragraphs could - if manipulated with - lead to biased results and maybe even biased conclusions. In a statistical analysis plan, emphasis should be on which analyses, comparisons, and statistical tests have been planned - given the objective of the study. The statistical methods section should include all the principal features of the proposed confirmatory analysis of the primary variable(s) and the way in which anticipated analysis problems will be handled. In the case of exploratory trials this section could describe more general principles and directions.

The set of participants whose data are to be included in the main analyses should be defined in the statistical section of the protocol. If there are any planned reasons, for excluding from analysis participants for whom data are available, these should be described. Some trials use terminology for these different scenarios (Analysis sets), such as: (#1) Full analysis set: The set of participants that is as close as possible to the ideal implied by the ITT principle. It is derived from the set of all randomized subjects by minimal and justified elimination of subjects; and, (#2) Per protocol set (valid cases, efficacy sample, evaluable participants sample): The set of data generated by the subset of participants who complied with the protocol sufficiently to ensure that these data would be likely to exhibit the effects of treatment, according to the underlying scientific model. In general

the ITT population (full analysis set) should be considered for all primary analyses. However, it is advantageous to demonstrate a lack of sensitivity of the principal trial results to alternative choices of the set of subjects analyzed.

When writing your statistical analysis plan, realize that data can be analyzed in many ways, although some of which may not be strictly appropriate in the particular situation. It is essential to specify which statistical procedure will be used for each analysis (given your objective and anticipated imaginary data set). Later in the full report (article) further clarification may be necessary in the results section, but this should never be in conflict with the analyses proposed in the protocol. When considering how elaborate such a paragraph needs to be, the principle is to describe statistical methods with enough detail to enable a knowledgeable reader with access to the original data¹⁷ to verify the reported results.

Most trial objectives result in statistical analyses yielding estimates of the treatment effect, which is a contrast between the outcomes in the comparison groups. These group contrasts should be followed by a confidence interval (usually 95% CI) for the estimated effect, which indicates a central range of uncertainty for the true treatment effect. Study findings can also be assessed in terms of their “statistical significance”. The “p-value” represents the probability that the observed data could have arisen by chance when the interventions did not truly differ. Which is why “very small” p-values (e.g., $p < 0.0001$) indicate that it is highly unlikely that the interventions on trial are equally good¹⁸; actual p-values (for example, $p = 0.031$) are strongly preferable to imprecise threshold reports such as $p < 0.05$.

Finally, standard methods of analysis assume that the data are “independent.” For controlled trials, this usually means that there is one observation per participant. Treating multiple observations from one participant as independent data is a serious error; such data are produced when outcomes can be measured on different parts of the body, as in dentistry or rheumatology.^{6,19} Data analysis should be based on counting each participant once (e.g., as a single change from baseline) or should be done by using more complex statistical procedures.^{20,21}

RESULTS AND OUTLINE

When you are finally done with conducting the procedures involved in your randomized trial, it is time to celebrate. You now have a very valuable database, where all your pre-specified objectives can be explored. Remember if you - or a junior colleague you are supervising - feel that it is too troublesome to write rigorous protocols, that there is a clear advantage at this stage: you have already written the first half of your article – you simply need to change the future tense into past tense. If you worry about what a good logical outline for the results section of such a trial could look like, we recommend that you scrutinize the examples given in the CONSORT statement,⁶ or follow other published examples with a rigorous reporting approach.²² Usually the following format is obvious: Figure 1: Trial flow diagram; Table 1: Baseline characteristics of participants randomized; Figure 2: Graphical display of the key findings from for the primary analysis and primary outcome (means or proportions with standard errors and 95% confidence interval); Table 2: Change from baseline at endpoint, for all the variables assessed in the study, both primary and secondary (means or proportions with standard errors and 95% confidence interval).

ACKNOWLEDGEMENTS AND FUNDING

This study was supported by unrestricted grants from The Oak Foundation.

REFERENCES

1. Altman DG, Goodman SN, Schroter S. How statistical expertise is used in medical research. *JAMA*. 2002;287:2817-2820.
2. Pocock SJ, Hughes MD, Lee RJ. Statistical problems in the reporting of clinical trials. A survey of three medical journals. *N Engl J Med*. 1987;317:426-432.
3. Dwan K, Altman DG, Arnaiz JA et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PLoS One*. 2008;3:e3081.
4. Ioannidis JP. The importance of potential studies that have not existed and registration of observational data sets. *JAMA*. 2012;308:575-576.
5. Altman DG, Simera I, Hoey J, Moher D, Schulz K. EQUATOR: reporting guidelines for health research. *Lancet*. 2008;371:1149-1150.
6. Moher D, Hopewell S, Schulz KF et al. CONSORT 2010 explanation and elaboration: updated guidelines

-
- for reporting parallel group randomised trials. *BMJ*. 2010;340:c869.
7. Vandenbroucke JP, von EE, Altman DG et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *Ann Intern Med*. 2007;147:W163-W194.
 8. Mamdani M, Sykora K, Li P et al. Reader's guide to critical appraisal of cohort studies: 2. Assessing potential for confounding. *BMJ*. 2005;330:960-962.
 9. Normand SL, Sykora K, Li P, Mamdani M, Rochon PA, Anderson GM. Readers guide to critical appraisal of cohort studies: 3. Analytical strategies to reduce confounding. *BMJ*. 2005;330:1021-1023.
 10. Rochon PA, Gurwitz JH, Sykora K et al. Reader's guide to critical appraisal of cohort studies: 1. Role and design. *BMJ*. 2005;330:895-897.
 11. Lewis JA. Statistical principles for clinical trials (ICH E9): an introductory note on an international guideline. *Stat Med*. 1999;18:1903-1942.
 12. Nuesch E, Reichenbach S, Trelle S et al. The importance of allocation concealment and patient blinding in osteoarthritis trials: a meta-epidemiologic study. *Arthritis Rheum*. 2009;61:1633-1641.
 13. Altman DG, Bland JM. Statistics notes: variables and parameters. *BMJ*. 1999;318:1667.
 14. Nuesch E, Trelle S, Reichenbach S et al. The effects of excluding patients from the analysis in randomised controlled trials: meta-epidemiological study. *BMJ*. 2009;339:b3244.
 15. Schulz KF, Grimes DA. Sample size calculations in randomised trials: mandatory and mystical. *Lancet*. 2005;365:1348-1353.
 16. Higgins JP, Altman DG, Gotzsche PC et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*. 2011;343:d5928.
 17. Hrynaszkiwicz I, Norton ML, Vickers AJ, Altman DG. Preparing raw clinical data for publication: guidance for journal editors, authors, and peer reviewers. *BMJ*. 2010;340:c181.
 18. Sterne JAC, Smith GD. Sifting the evidence - what's wrong with significance tests? *BMJ*. 2001;322:226-231.
 19. Ranstam J. Repeated measurements, bilateral observations and pseudoreplicates, why does it matter? *Osteoarthritis Cartilage*. 2012;20:473-475.
 20. Ellegaard K, Torp-Pedersen S, Christensen R et al. Feasibility of a standardized ultrasound examination in patients with rheumatoid arthritis: a quality improvement among rheumatologists cohort. *BMC Musculoskelet Disord*. 2012;13:35.
 21. Boesen AP, Boesen MI, Torp-Pedersen S et al. Associations between abnormal ultrasound color Doppler measures and tendon pain symptoms in badminton players during a season: a prospective cohort study. *Am J Sports Med*. 2012;40:548-555.
 22. Bliddal H, Leeds AR, Stigsgaard L, Astrup A, Christensen R. Weight loss as treatment for knee osteoarthritis symptoms in obese patients: 1-year results from a randomised controlled trial. *Ann Rheum Dis*. 2011;70:1798-1803.