

PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Reporting quality of clinical trial protocols: a repeated cross sectional study about the Adherence to SPIrit Recommendations in Switzerland, CA Canada, and GERmany (ASPIRE-SCAGE)
AUTHORS	Gryaznov, Dmitry; von Niederhäusern, Belinda; Speich, Benjamin; Kasenda, Benjamin; Ojeda-Ruiz, Elena; Blümle, Anette; Schandelmaier, Stefan; Mertz, Dominik; Odutayo, Ayodele; Tomonaga, Yuki; Amstutz, Alain; Pauli-Magnus, Christiane; Gloy, Viktoria; Lohner, Szimonetta; Bischoff, Karin; Wollmann, Katharina; Rehner, Laura; Meerpohl, Joerg; Nordmann, Alain; Klatte, Katharina; Ghosh, Nilabh; Taji Heravi, Ala; Wong, Jaqueline; Chow, Ngai; Hong, Patrick; McCord, Kimberly; Sricharoenchai, Sirintip; Busse, Jason; Agarwal, Arnav; Saccilotto, Ramon; Schwenkglenks, Matthias; Moffa, Giusi; Hemkens, Lars; Hopewell, Sally; Von Elm, Erik; Briel, Matthias

VERSION 1 – REVIEW

REVIEWER	Mew, Emma Yale University School of Public Health, Chronic Disease Epidemiology
REVIEW RETURNED	26-Jul-2021

GENERAL COMMENTS	<p>Thank you for the opportunity to read this interesting and important manuscript. This manuscript describes a cross sectional pre-post design that aimed to assess the completeness of clinical trial protocol reporting immediately before (2012) and three years after (2016) the release of the SPIRIT reporting guideline in 2013. The manuscript used an international sampling approach and assessed the reporting of nearly all clinical trial protocols in Switzerland and a small convenience sample of clinical trial protocols at study sites in Germany and Canada. The authors analyzed the results using two primary regression modeling approaches – beta regression and logistic regression -- and also conducted sensitivity analyses using methods outlined a priori in a previously published protocol.</p> <p>This is a well-written, clear, and polished manuscript that will add value to the field of meta-research and evaluation of reporting guidelines. The methods appear rigorous and sufficiently detailed. Based on this, I recommend this manuscript be accepted with minor revisions. I have included a few minor suggestions that the authors might consider to strengthen the work (below). My primary suggestion would be to provide more information and discussion on risk for bias and generalizability.</p> <p>1. Strengths and limitations:</p>
-------------------------	--

	<p>a. Line 26: Please highlight that the international sample sizes were considerably smaller than the Swiss sample. One suggestion could be to modify the text to “small convenience sample”.</p> <p>2. Introduction:</p> <p>a. Line 5-12: If space permits, expand on why incomplete clinical trial protocol reporting is problematic to further underscore the importance of this work to a lay reader. For example, why are complete protocols are needed to appraise quality? Why would they compromise patient safety?</p> <p>3. Methods:</p> <p>a. Line 17: Is there a reason this study restricted to clinical trials and not RCTs generally?</p> <p>b. Was screening for eligibility also done independently and in duplicate?</p> <p>c. Is there a reason the sample size term was coded in 1000 increments instead of the raw value?</p> <p>d. What test was used to test for a significant increase from 2012 to 2016? For example, did you use a Wald/X2 test based on the beta term in the model, or did you perform separate likelihood ratio tests (or an alternative method)?</p> <p>4. Results:</p> <p>a. Would it be possible to state the exact sample sizes for Canada and Germany in the main body of the manuscript?</p> <p>b. The manuscript stated that most trials were in the areas of oncology or CV medicine. Would it be possible to provide some quantified information towards the breakdown of disease disciplines for the included protocols?</p> <p>c. P17 L3-15: Please state the results of the statistical tests performed to assess statistically significant differences between years (and exact p-values).</p> <p>d. P18 L5: Would it be possible to briefly explain why a multivariable beta regression model approach was used, and in what instances this approach was used over a logistic approach? Providing model equations and terms as a supplementary material might be an approach to consider to better help guide the reader.</p> <p>e. Page 15 Line 4-10: Would it be possible to include the p-values and odds ratios in the text to help guide the reader?</p> <p>f. Page 17 line 5: What percentage of RECs participated in Switzerland?</p> <p>g. Figure 2: Would it be possible to interpret the ORs so the reader is clear towards the reference group, etc.? Were these ORs generated using adjusted or unadjusted models?</p> <p>5. Discussion:</p> <p>a. Would it be possible to further expand on risk of bias: (1) Would the 15% of protocols not evaluated in duplicate (line 47) be at potential risk for bias? For example, would it be possible to comment on whether this a random sample of all records, or records from just one country, etc.? (2) Is there possibility of risk of bias from a lack of blinding of assessors to other information in the protocol (for example, a possibility for bias between industry-sponsored over investigator-sponsored protocols)?</p> <p>b. Please expand further on generalizability: (1) Did any of the RECs sampled in 2016 explicitly endorse SPIRIT? If so, how would this impact generalizability?; and (2) Would it be possible to comment on the generalizability of results to disease areas outside of oncology and cardiovascular disease?</p>
--	--

	<p>c. Based on Figure 1, it is interesting that no protocols sampled in 2016 met all SPIRIT items. Would you have an interpretation of this? I am curious if your team feels this might reflect a feasibility problem with the current SPIRIT checklist?</p> <p>6. Supplementary materials: a. Figure 2: What do you mean by tertile? Would it be possible to provide more context as to why tertiles were meaningful to present in this figure?</p>
--	---

REVIEWER	Cheng, Chao Sun Yat-sen University First Affiliated Hospital, thoracic surgery
REVIEW RETURNED	28-Nov-2021

GENERAL COMMENTS	<p>The authors reported and compared the proportion of trial protocols reporting individual SPIRIT items per protocol in 2012 and 2016, and built regression models to explore factors associated with adherence to SPIRIT. Based on the findings, they concluded after publication of the SPIRIT checklist, investigator sponsored protocols improved to the level of industry-sponsored protocols, which did not improve. In addition to the above results, I'd like to see the impact of the change on the quality of investigator-sponsored RCT in the results and/or discussion part.</p>
-------------------------	--

VERSION 1 – AUTHOR RESPONSE

B) Comments from Reviewer 1 (Ms. Emma Mew)

Thank you for the opportunity to read this interesting and important manuscript. This manuscript describes a cross-sectional pre-post design that aimed to assess the completeness of clinical trial protocol reporting immediately before (2012) and three years after (2016) the release of the SPIRIT reporting guideline in 2013. The manuscript used an international sampling approach and assessed the reporting of nearly all clinical trial protocols in Switzerland and a small convenience sample of clinical trial protocols at study sites in Germany and Canada. The authors analyzed the results using two primary regression modeling approaches – beta regression and logistic regression -- and also conducted sensitivity analyses using methods outlined a priori in a previously published protocol.

This is a well-written, clear, and polished manuscript that will add value to the field of meta-research and evaluation of reporting guidelines. The methods appear rigorous and sufficiently detailed. Based on this, I recommend this manuscript be accepted with minor revisions. I have included a few minor suggestions that the authors might consider to strengthen the work (below). My primary suggestion would be to provide more information and discussion on risk for bias and generalizability.

Author reply: We thank the reviewer for the positive feedback and suggestions to further improve the manuscript.

Abstract

1. Strengths and limitations:

a. Line 26: Please highlight that the international sample sizes were considerably smaller than the Swiss sample. One suggestion could be to modify the text to “small convenience sample”.

Author reply: We revised the strengths and limitations section after the abstract as requested by the Editor (see above). We fully agree with your suggestion and rephrased the corresponding text as follows:

“The sample of trial protocols from Switzerland (n=397) was much larger than the sample from Germany (n=75) or Canada (n=77).”

2. Introduction

a. Line 5-12: If space permits, expand on why incomplete clinical trial protocol reporting is problematic to further underscore the importance of this work to a lay reader. For example, why are complete protocols are needed to appraise quality? Why would they compromise patient safety?

Author reply: Thank you for this suggestion which we followed. We added the following sentences to the Introduction section:

“With incomplete protocols reviewers typically cannot distinguish between the use of inappropriate methodology and the non-reporting of appropriate methodology. In addition, if details about the application of the trial intervention or situations with un-blinding of trial participants are lacking, the resulting uncertainty with treating clinicians may compromise the safety of trial participants.”

3. Methods:

a. Line 17: Is there a reason this study restricted to clinical trials and not RCTs generally?

Author reply: We define an RCT as a prospective study in which patients, or groups of patients, are assigned at random to one or more interventions to evaluate their effect on health outcomes. We excluded trials enrolling healthy volunteers (e.g., pharmacokinetic studies, training interventions in sport science), economic evaluations, animal studies, studies based on tissue samples, observational studies, studies involving only qualitative methods, and studies with a quasi-random method of allocation, as the SPIRIT guidelines were not strictly intended for these trials and also the same level of rigor is not to be expected in their protocols. We felt that the term “randomised clinical trial” would capture this aspect better than the term “randomised controlled trial”.

b. Was screening for eligibility also done independently and in duplicate?

Author reply: Thank you for raising this point. Yes, screening for eligibility was done independently and in duplicate too. We added the following to the first paragraph of the Methods section:

“The eligibility of RCT protocols was assessed independently and in duplicate. Any disagreements were resolved by discussion and consensus.»

c. Is there a reason the sample size term was coded in 1000 increments instead of the raw value?

Author reply: Thank you for the opportunity to clarify this point. The term was coded sample size/1000 so that the estimates of the regression were approximately of the same decimal facilitating the readability of the results (e.g. if coded raw sample size $\exp(0.000021)$ would be 1.000021, we used increments of $1000 \exp(0,0210)$ resulting in 1.02122).

d. What test was used to test for a significant increase from 2012 to 2016? For example, did you use a Wald/ χ^2 test based on the beta term in the model, or did you perform separate likelihood ratio tests (or an alternative method)?

Author reply: In the beta regression model hypothesis testing was performed using approximations obtained from the asymptotic normality of the maximum likelihood estimator, as described by Ferrari S, Cribari-Neto F. Beta regression for modelling rates and proportions. J Appl Stat. 2004;31(7):799–

815, and using the outputs of the betareg function (<https://cran.r-project.org/web/packages/betareg/betareg.pdf>).

4. Results:

a. Would it be possible to state the exact sample sizes for Canada and Germany in the main body of the manuscript?

Author reply: We followed the suggestion of the Reviewer and include now the exact sample sizes for Canada and Germany in the first paragraph of the Results section.

b. The manuscript stated that most trials were in the areas of oncology or CV medicine. Would it be possible to provide some quantified information towards the breakdown of disease disciplines for the included protocols?

Author reply: Thank you for raising this point. We now include a breakdown of included trial protocols by medical discipline in the additional Supplementary Table 9.

c. P17 L3-15: Please state the results of the statistical tests performed to assess statistically significant differences between years (and exact p-values).

Author reply: In this section of the manuscript we simply report median proportions of reported SPIRIT-items for investigator-sponsored protocols and for industry-sponsored protocols, just descriptively as observed. Testing for statistical significance only occurred in multivariable regression models to minimize confounding. The results of our multivariable regression analyses are reported in the last two paragraphs of the Results section.

d. P18 L5: Would it be possible to briefly explain why a multivariable beta regression model approach was used, and in what instances this approach was used over a logistic approach? Providing model equations and terms as a supplementary material might be an approach to consider to better help guide the reader.

Author reply: As pre-specified in our protocol (Gryaznov et al. *Trials* 2020) and mentioned in the Methods section, multivariable beta regression was used for all primary statistical analyses. We used beta regression for all primary analyses, because it allowed us to directly model the proportion of SPIRIT items adhered to per protocol. This is not possible with logistic regression. Hierarchical logistic regression was used as a sensitivity analysis, because the aggregated proportion as a response (beta regression) did not allow us to capture the variability within each protocol. The hierarchical logistic regression model considers two levels: the "SPIRIT item level" and the "protocol level". The response is a binary variable indicating adherence to each SPIRIT item with clustering by protocol.

To clarify this better for the reader we added the following sentence to the Data analysis section of the Methods:

"Beta regression allowed us to directly model the proportion of SPIRIT items adhered to per protocol²¹, while hierarchical logistic regression allowed us to capture the variability within protocols."

In addition, we clarified that the results from multivariable regression reported in the Results section under the subheading "Multivariable regression analysis" come from beta regression. In the last sentence of the Results section we now state the following:

"Sensitivity analyses using hierarchical logistic regression instead of beta regression confirmed all results."

e. Page 15 Line 4-10: Would it be possible to include the p-values and odds ratios in the text to help guide the reader?

Author reply: We now added odds ratios and p-values to the revised main text as requested.

f. Page 17 line 5: What percentage of RECs participated in Switzerland?

Author reply: All RECs in Switzerland participated in our study (i.e. 100%). To make this clearer we dropped the word “practically” in the respective sentence of the Discussion section.

g. Figure 2: Would it be possible to interpret the ORs so the reader is clear towards the reference group, etc.? Were these ORs generated using adjusted or unadjusted models?

Author reply: We added more information to Figure 2 and its legend in order to better clarify for the reader reference groups and the fact that ORs were derived from adjusted (multivariable) models.

5. Discussion:

a. Would it be possible to further expand on risk of bias: (1) Would the 15% of protocols not evaluated in duplicate (line 47) be at potential risk for bias? For example, would it be possible to comment on whether this is a random sample of all records, or records from just one country, etc.? (2) Is there possibility of risk of bias from a lack of blinding of assessors to other information in the protocol (for example, a possibility for bias between industry-sponsored over investigator-sponsored protocols)?

Author reply: Thank you for the opportunity to clarify this point. We state in the Methods section of the revised manuscript that “researchers trained in trial methodology completed a calibration process to improve reliability, and then extracted relevant data from RCT protocols independently and in duplicate, including whether individual SPIRIT items were reported.¹⁹ Disagreements were resolved by discussion. Due to limited resources 15% of included protocols were extracted by a single researcher (having extracted at least 100 RCT protocols in duplicate). All researchers extracting data from RCT protocols signed confidentiality agreements and the final database contained only coded data.” The 15% of protocols not evaluated in duplicate were from different RECs in Switzerland. Since a large majority of all included protocols were from Switzerland and since protocols not evaluated in duplicate were handled by one of two very experienced data extractors only, we do not feel that a relevant increase in the risk of bias is plausible. We added the following to the limitations section of the revised manuscript:

“Fourth, 15% of included protocols were not evaluated in duplicate which could have increased the risk of bias in our study. However, these protocols were from different RECs in Switzerland and they were handled by one of the two most experienced data extractors only, so we feel that a relevant increase in the risk of bias is unlikely.”

Considering the large number of data extractors and considering the fact that we did not conduct interim analyses in our study, we do not see plausible reasons why lack of blinding of SPIRIT assessors to various trial characteristics should have biased our results. In addition, blinding of SPIRIT assessors to various trial characteristics would not have been possible given the content of the 33 SPIRIT items.

b. Please expand further on generalizability: (1) Did any of the RECs sampled in 2016 explicitly endorse SPIRIT? If so, how would this impact generalizability?; and (2) Would it be possible to comment on the generalizability of results to disease areas outside of oncology and cardiovascular disease?

Author reply: Thank you for raising this point. We added the following to the limitations section of the revised manuscript:

“Fifth, we are not aware of the fact that any of the participating RECs explicitly endorsed SPIRIT guidance, however, in Switzerland a new protocol template provided by swissethics became available which was influenced by SPIRIT impacting the generalisability of our results. In addition, it remains unclear to what extent our findings can be extrapolated to trial protocols from middle- or low-income countries and to protocols from medical disciplines underrepresented in our sample (e.g. dentistry or geriatrics; Supplementary Table 9).”

In addition, we rephrased the last sentence of the paragraph about the strengths as follows:

“The fact that all Swiss RECs participated in this study strengthens the representativeness of our data for Switzerland and the additional inclusion of a German and a Canadian REC allowed for an international comparison to some extent.”

c. Based on Figure 1, it is interesting that no protocols sampled in 2016 met all SPIRIT items. Would you have an interpretation of this? I am curious if your team feels this might reflect a feasibility problem with the current SPIRIT checklist?

Author reply: Thank you for this comment. We agree with the reviewer that this may indicate some feasibility problem with the current SPIRIT checklist. Further experiences from other groups applying the SPIRIT checklist to a larger number of RCT protocols would be relevant to discuss and optimize the content and most adequate operationalization of SPIRIT items. We are very happy to contribute to the further development of SPIRIT guidance. However, we feel that this aspect goes beyond the scope of the present manuscript.

6. Supplementary materials:

a. Figure 2: What do you mean by tertile? Would it be possible to provide more context as to why tertiles were meaningful to present in this figure?

Author reply: Under tertiles we understand the representation of the calendar year divided into three equal parts (4 months each). We wanted to investigate for any potential trends over time within 2012 and 2016 as an indication for a potential overall time trend (e.g. continuous improvement over time that would also result in a difference between protocol reporting quality in 2012 and 2016). Tertiles appeared appropriate, because this allowed for a substantial number of observations (trials protocols) to estimate the SPIRIT adherence of protocols per time period (tertile).

C) Comments from Reviewer 2 (Dr. Chao Cheng)

The authors reported and compared the proportion of trial protocols reporting individual SPIRIT items per protocol in 2012 and 2016, and built regression models to explore factors associated with adherence to SPIRIT. Based on the findings, they concluded after publication of the SPIRIT checklist, investigator sponsored protocols improved to the level of industry-sponsored protocols, which did not improve. In addition to the above results, I'd like to see the impact of the change on the quality of investigator-sponsored RCT in the results and/or discussion part.

Author reply: Thank you for your comment. We generally feel that the potential impact of results (in our case the improvement of the quality of investigator-sponsored RCT protocols) should not be reported/discussed in the Results section of a manuscript. In the Discussion section we addressed the impact of the change in the quality of investigator-sponsored RCTs in the revised manuscript under the subheading “Implications” as follows:

“Whether there is indeed an association between better reported or more comprehensive RCT protocols and better methodology, successful trial conduct, and/or publication of RCTs remains to be established. Based on the RCT sample of this study, we will examine the relationship between completeness of RCT protocols and risks for premature discontinuation or non-publication of RCTs as well as potential improvements between 2012 and 2016 in terms of fewer trial discontinuations and non-publications particularly for investigator-sponsored RCTs in subsequent investigations 19.”

Additional response to request from 18th of Jan 2022:

1. Author's response:

- Please provide a point-by-point response to the Editor's comments and reviewer's comments.

Uploaded and entered in form.

2. Required Supplementary format:

- Please re-upload your Supplementary files in PDF format.

Clean and marked copy of the Supplementary files uploaded in PDF format.

3. Author's name mismatch:

- The author "Jacqueline Wong" in your main document is registered as "Jaqueline Wong" in ScholarOne. Please ensure that the author has same registered name.

Jacqueline Wong is the correct name. She is registered in orchid under the number 0000-0002-7583-8137, e-mail: wongj37@mcmaster.ca. I could not add her via the ORCHID ID in any way in the online form. Please advise how to correct this? Should I ask the author to change her registration in ScholarOne?

4. Incomplete contributorship statement:

- Please provide a more detailed contributorship statement. It needs to mention all the names/initials of authors along with their specific contribution/participation for the article. *Katharina Wollmann and Giusi Moffa not mentioned in contributorship statement

This mistake was corrected, the statement amended, missing contributors added.