

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)	Development and Psychometric Evaluation of the Implementation Science Research Project Appraisal Criteria (ImpResPAC) tool: A Study Protocol
AUTHORS	SWEETNAM, CHLOE; Goulding, L; Davis, Rachel; Khadjesari, Zarnie; Boaz, Annette; Healey, Andy; Sevdalis, Nick; Bakolis, Ioannis; Hull, Louise

VERSION 1 – REVIEW

REVIEWER	Marie-Therese Schultes University of Zurich, Institute for Implementation Science in Health Care
REVIEW RETURNED	25-Mar-2022

GENERAL COMMENTS	<p>Thank you for the opportunity to review the study protocol "Development and Psychometric Evaluation of the Implementation Science Research Project Appraisal Criteria (ImpResPAC) tool". I am convinced that the ImpResPAC tool will provide a valuable contribution to the development of implementation science, which currently lacks standardized appraisal criteria for research projects. In their study protocol, the authors thoroughly describe methods that will certainly contribute to the quality of the tool. Please find below some suggestions for the further development of the study protocol:</p> <p>Introduction In both the introduction and discussion section, the authors describe that the tool will contribute to capacity building and be of use for educators in implementation science. However, there is a high variability in implementation science training and education in terms of trainees' backgrounds and competence levels and it seems unlikely that student projects in most contexts could be rated on all the tool's domains. Accordingly, I would encourage the authors to elaborate further on how they expect the tool to contribute to implementation science education. Moreover, Part B of the attached survey lists researchers and practitioners as potential users of the tool, but these stakeholders are hardly mentioned in the manuscript. It would be helpful to also have some information on how the tool could be beneficial for self-assessments of researchers' and practitioners' own projects.</p> <p>On p.7, describing the initial study on developing the tool, the authors state that "Intra-class correlation coefficient (ICC) tests indicated excellent inter-rater reliability (IRR)". Please provide the respective values in the text or cite the original reference again. Also, could the authors provide an explanation why the previous study focused on exactly these five ImpResPAC domains?</p>
-------------------------	--

	<p>On p.8, regarding differences between ImpResPAC and INSPECT, the authors state that “INSPECT primarily focuses on fundability whereas ImpResPAC focuses on conceptual and methodological quality of implementation research”. Still, funders are mentioned as one of the main stakeholder groups to make use of ImpResPAC and conceptual and methodological quality of research are important funding criteria. Hence, I was wondering whether the authors could provide further criteria that distinguish the two tools. In the discussion section, it is mentioned that ImpResPAC includes more recent methodological guidelines than INSPECT. Here, it would be helpful if the authors could list examples of such guidelines.</p> <p>Methods and analysis On p.9, the initial ImpResPAC development study is mentioned again, this time called a “dissertation study”, which seems not necessary to mention explicitly.</p> <p>Regarding the tool’s content validation and refinement, it would be helpful if the authors provided information on the development of the specific items associated with the domains. Are there example items that can be shared in the protocol? This would also be useful for the clarity of the expert survey questions reviewing these items.</p> <p>Concerning the expert survey, would it be possible to develop an online survey with filter questions, where experts can indicate their domains of expertise and are then led to the corresponding survey questions? Also, the “Overview of ImpResPAC” and “ImpResPAC user instructions” contain some overlap in text that could be reduced.</p> <p>On p.9, please also indicate whether in the second round of reviews, experts are going to review their “own” domains again or the whole tool.</p> <p>P.10: The abbreviations “IF” and “D&I” are only used once and therefore don’t need to be introduced.</p> <p>On p.11, please indicate in the beginning of the assessment description of convergent validity whether the whole ImpResPAC or just certain dimensions will be correlated with the INSPECT scale.</p> <p>I found table 1 very helpful for a comparison of the dimensions of the two tools. However, I was surprised that there is no overlap between the INSPECT dimension “policy/funding environment” and the ImpResPAC dimension “determinants of implementation contextual factors”. Maybe the authors can provide a brief explanation for that.</p> <p>Discussion On p.17, it is stated that the tool should help funders “differentiate high- and low-quality implementation research”. Since the tool contains several different domains to assess quality, I was wondering whether a more differentiated view on research quality would be more appropriate to describe the tool’s benefits, e.g., that it helps funders to “identify specific areas of improvement”.</p>
--	---

REVIEWER	Toto Gronlund NIHR, James Lind Alliance
REVIEW RETURNED	16-Apr-2022

GENERAL COMMENTS	Thank you for submitting this protocol. It is clearly written, and has
-------------------------	--

	<p>potential to be a worthwhile extension of the ImpRes guide.</p> <p>I have a couple of comments you may wish to consider. As lay reviewer, I was surprised that there is no public or patient involvement in the design of the study, or indeed intention to involve members of the public or patients as experts in any of the stages of the study. It would seem pertinent to involve them, especially given the ImpResPAC tool includes a PPI domain. PPI experts should represent a significant proportion of your target expert advisory group, as well as be included in the surveys.</p> <p>The proposal lacks detail on how the content of the existing INSPECT tool and the proposed ImpResPAC tool differ, other than at domain level. It is difficult, therefore, to see what are the assumed conceptual and methodological weaknesses INSPECT, which this proposal intends to review and enhance with ImpResPAC.</p> <p>Without knowing the content of either too, from my perspective, a head-to-head comparison of the "results" of using the two different tools on planned the case set of studies (in Stage 3), would be an important enhancement to this proposal in terms of its methodology. The proposal states that the application of the ImpResPAC tool in stage 3 would be done by the researchers. Is there opportunity to include the other possible end users in Stage 3, (categories as per survey questions) namely those reviewing grant applications, practitioners, and educators. I would also suggest PPI reps here. Perhaps members of the expert advisory group could pilot / test the tool?</p> <p>The comparison of domains of INSPECT and the proposed ImpResPAC is helpful, but as mentioned before, without much detail (Table 1). I am guessing this is because it can be assumed that people will know the content of both INSPECT and the ImpRes guide. Nevertheless, the comparison seems a bit arbitrary, and may need just to be reviewed. Will this potentially change during Stage 1 of the study? In other words, should this be presented as an initial mapping, to be refined, by Stage 1 and then by survey part A? Also, the comparison in Table 1 does seem to suggest that INSPECT has several specific domains about context: care and quality gap; setting's readiness; policy environment. On the other hand the proposed ImpResPAC appears to roll these into one domain - contextual factors - albeit in the ImpRes guide this is about a wide range of contextual factors.</p> <p>Some minor points: It is not clear when it is intended to run the study. Also, data ranges should be considered for the inclusion and exclusion criteria, or an explanation of why open dates are being used.</p> <p>What is the planned reach of the survey, in terms of range of stakeholders, countries, numbers ?</p> <p>I am not able to comment on the planned statistical analysis.</p>
--	---

REVIEWER	Alicia Bungler The Ohio State University
REVIEW RETURNED	27-Apr-2022

GENERAL COMMENTS	Thank you for the opportunity to review this study protocol. This sequential mixed-methods study is designed to develop a
-------------------------	---

	<p>quantitative tool for appraising the quality of implementation studies. I believe this tool could benefit the field. My comments below are intended to help strengthen the presentation, and clarify the potential contributions (and limitations) of this work.</p> <p>Authors note that a lack of practical guidance likely contributes to poor quality implementation research. I'm wondering if there are examples or evidence that suggests that offering practical tools for study design improves research quality (if not in implementation, then in other fields). This would bolster the justification for this type of tool. For instance, there are many tools and resources for program evaluation – do we know if this actually improves the quality of evaluations? In implementation, we've seen many journals require that authors adhere to set reporting guidelines. Does this actually improve quality (at least of dissemination)?</p> <p>I also think it would be helpful to define what “high quality” implementation research is, from who’s perspective, and whether there is consensus (or not) in the field. Given the broad range of studies I'd imagine that definitions vary widely across stakeholder groups. For instance, my funders might expect methodological rigor and somewhat controlled designs, while my community partners value applicability/utility. Does the inclusion of “experts” as defined potentially exclude perspectives that could be useful? How might this limit the applicability of your tool? And does this reify dominant standards of research?</p> <p>The methods are staged thoughtfully. The description of Stage 1 suggests that psychometric testing will be conducted at this point of the study, but I think that actually happens in Stage 3. This could be clarified further. Additional detail about how the tool will be refined/developed would be useful.</p> <p>Finally, while I think practical tools for improving rigor could be extremely useful in a young and growing field, it might be worthwhile to consider the potential unintended consequences of such a tool. Does formalizing these quality standards restrict or disincentivize innovation? Disadvantage particular types of topics, forms of research, designs, etc? Prevent the introduction/application of different theoretical perspectives? Does it allow for evolution of the field and what we consider rigorous? If so, are these tradeoffs acceptable for the field?</p>
--	--

VERSION 1 – AUTHOR RESPONSE

Reviewer 1

I am convinced that the ImpResPAC tool will provide a valuable contribution to the development of implementation science, which currently lacks standardized appraisal criteria for research projects. In their study protocol, the authors thoroughly describe methods that will certainly contribute to the quality of the tool. Please find below some suggestions for the further development of the study protocol

We thank the reviewer for their thoughtful and thorough review of our manuscript. We are delighted that the reviewer appreciates that ImpResPAC tool will provide a valuable contribution to the development of implementation science, which currently lacks standardized appraisal criteria for research projects.

In what follows, we respond to the reviewer’s detailed review of our manuscript and suggestions for improving the quality of writing.

I would encourage the authors to elaborate further on how they expect the tool to contribute to implementation science education.

We thank the review for highlighting that the contribution of ImpResPAC to implementation science education was not explicit before. We envisage that educators will be able to use ImpResPAC to appraise implementation projects submitted by students/learners. We acknowledge that implementation science curricula may be different across courses, but the core concepts are generic, and not all ImpResPAC domains will apply to all projects.

We have now updated the manuscript to reflect how we envisage ImpResPAC to contribute to implementation science education.

The sentence (line 363-366) now reads:

“In addition, we also envisage that ImpResPAC will be useful to educators that are tasked with appraising implementation projects submitted by students/learners, especially in educational settings where the ImpRes tool and guide informed the curriculum.”

Moreover, Part B of the attached survey lists researchers and practitioners as potential users of the tool, but these stakeholders are hardly mentioned in the manuscript.

It would be helpful to also have some information on how the tool could be beneficial for self-assessments of researchers’ and practitioners’ own projects.

We recognize that it was not clear enough who we envisage the primary user of ImpResPAC to be.

We propose that ImpResPAC will be used by implementation stakeholders in the following way:

1. Funders: As the primary users, use ImpResPAC to appraise project proposals.
2. Educators: Use ImpResPAC to appraise projects submitted by students/learners. We acknowledge that implementation science curricula may be different, but the core concepts are generic, and not all ImpResPAC domains will apply to all projects.
3. Researchers, practitioners and students: We plan to explore whether another potential application of ImpResPAC would be for implementation researchers, practitioners and students/learners to use ImpResPAC as a quality assurance step, to self-assess a funding application or implementation project, prior to submission.

We have updated the manuscript to make this clearer. The sentences (line 360-369) now read:

“We envisage ImpResPAC primarily being used by funding bodies as a standardized and transparent method to differentiate high and low-quality implementation research and prospectively identify areas for improvement before funding decisions are made. In addition, we also envisage that ImpResPAC will be useful to educators that are tasked with appraising implementation projects submitted by students/learners, especially in educational settings where the ImpRes tool and guide informed the curriculum. We plan to explore whether another potential application of ImpResPAC would be for implementation researchers, practitioners, and students/learners to use ImpResPAC as a quality assurance step, to self-assess a funding application or implementation project, prior to submission.”

On p.7, describing the initial study on developing the tool, the authors state that “Intra-class correlation coefficient (ICC) tests indicated excellent inter-rater reliability (IRR)”. Please provide the respective values in the text or cite the original reference again.

We have amended as suggested. The sentence (line 151-152) now reads:

“The resulting Intra-class correlation coefficient (ICC) was in the ‘excellent’ inter-rater reliability (IRR) range: ICC: 0.85 (23).”

Also, could the authors provide an explanation why the previous study focused on exactly these five ImpResPAC domains?

We appreciate that the reasons for only focusing on five ImpResPAC domains was not clear.

We have amended to include an explanation. Line 131-141, with new text in bold, now reads:

“Another example of prior efforts to quantify the quality of implementation research, by some of the authors of this paper (CS, LG, LH), reported the initial development of a quantitative appraisal tool, based on the ImpRes tool and supplementary guide (20) (21) as part of a master’s dissertation project. Due to time constraints and scope of the master’s dissertation project, this initial development work focused on five of the ten ImpRes domains: 1) Implementation research characteristics; 2) Implementation theories, frameworks, and models; 3) Determinants of implementation; 4) Implementation strategies; 5) Implementation outcomes. These domains were considered to be most relevant and specific to implementation research, whereas the other domains (e.g., service and patient outcome), while still relevant to implementation research, overlap over research types (e.g., effectiveness research).

On p.8, regarding differences between ImpResPAC and INSPECT, the authors state that “INSPECT primarily focuses on fundability whereas ImpResPAC focuses on conceptual and methodological quality of implementation research”. Still, funders are mentioned as one of the main stakeholder groups to make use of ImpResPAC and conceptual and methodological quality of research are important funding criteria. Hence, I was wondering whether the authors could provide further criteria that distinguish the two tools.

We acknowledge that the initial mapping of domains did not show granular detail to distinguish the two tools. Since stage 1 (mapping of ImpRes tool and guide to develop ImpResPAC domain items) is now complete, we have added a comparison of the ImpResPAC domain items and INSPECT element items. This comparison includes an initial mapping of the ImpRes tool and guide to develop ImpResPAC and is subject to change after survey part A results have been analysed by the ImpResPAC team.

For example, ‘team experience with setting, treatment, and implementation process’ is one of the ten elements of the INSPECT tool, however the ImpRes tool and supplementary guide, and consequently ImpResPAC, will not contain criteria measuring this domain as team experience is not a direct measure of conceptual or methodological quality of IS research.

Furthermore, the level of detail at which implementation research will be appraised using the two scoring systems will differ substantially. For example, INSPECT provides an overall appraisal of the measurement and analysis of IS research proposals, however the ImpRes guide, and consequently ImpResPAC, will contain three domains relating to measurement and analysis; 1) service and patient outcomes; 2) implementation outcomes; and 3) economic evaluation, providing a much more detailed and focused appraisal of the outcomes typically assessed in implementation research.

We have updated the protocol to include this more detailed comparison of the ImpResPAC tool and the INSPECT tool to provide further criteria distinguishing the two tools. See supplementary material 2. Line 407-410 now reads:

The initial mapping of the ImpRes tool and supplementation guide to develop the ImpResPAC tool (stage 1) and a detailed comparison of ImpResPAC tool domain items (initial mapping) and the INSPECT tool element items can be found in supplementary material (additional file 2).

In the discussion section, it is mentioned that ImpResPAC includes more recent methodological guidelines than INSPECT. Here, it would be helpful if the authors could list examples of such guidelines.

We have amended as suggested and included examples of these guidelines, cited in bold (line 418):

As such, ImpResPAC will include and operationalize key methodological guidelines and recommendations that simply did not exist nearly a decade ago (1) (8) (10) (33) (34) (35) (36) (37) (38) (39). ImpResPAC will operationalize, for example, the key methodological and conceptual guidelines and recommendations that have been described in the ImpRes tool and guide, as well as guidelines suggested by our international expert advisory panel, and key literature published since the development of the ImpRes tool and guide.

On p.9, the initial ImpResPAC development study is mentioned again, this time called a “dissertation study”, which seems not necessary to mention explicitly.

We have amended as suggested and removed the word ‘dissertation’ (line 193-194). The sentence now reads:

As part of a previous study, five of the ImpResPAC domains were developed and inter-rater reliability was assessed (20).

Regarding the tool’s content validation and refinement, it would be helpful if the authors provided information on the development of the specific items associated with the domains. Are there example items that can be shared in the protocol? This would also be useful for the clarity of the expert survey questions reviewing these items.

We have updated the protocol to include this more detailed comparison of the ImpResPAC tool and the INSPECT tool to provide further criteria distinguishing the two tools. For example, ‘team experience with setting, treatment, and implementation process’ is one of the ten elements of the INSPECT tool, however the ImpRes tool and supplementary guide, and consequently ImpResPAC, will not contain criteria measuring this domain as team experience is not a direct measure of conceptual or methodological quality of IS research.

Furthermore, the level of detail at which implementation research will be appraised using the two scoring systems will differ substantially. For example, INSPECT provides an overall appraisal of the measurement and analysis of IS research proposals, however the ImpRes guide, and consequently ImpResPAC, will contain three domains relating to measurement and analysis; 1) service and patient outcomes; 2) implementation outcomes; and 3) economic evaluation, providing a much more detailed and focused appraisal of the outcomes typically assessed in implementation research. See supplementary material 2 for a detailed comparison.

We have updated the protocol to include this more detailed comparison of the ImpResPAC tool and the INSPECT tool to provide further criteria distinguishing the two tools. See supplementary material 2. Line 407-410 now reads:

The initial mapping of the ImpRes tool and supplementation guide to develop the ImpResPAC tool (stage 1) and a detailed comparison of ImpResPAC tool domain items (initial mapping) and the INSPECT tool element items can be found in supplementary material (additional file 2).

Concerning the expert survey, would it be possible to develop an online survey with filter questions, where experts can indicate their domains of expertise and are then led to the corresponding survey questions?

We thank the reviewers for their useful suggestion regarding incorporating filter questions. Although data collection for survey A is now complete it is possible to incorporate filter questions in survey B to streamline data collection.

Also, the “Overview of ImpResPAC” and “ImpResPAC user instructions” contain some overlap in text that could be reduced.

The ImpResPAC user instructions have been reviewed by experts (as part of Survey A, now complete) and will be refined by the research/development team, also as part of stage 2 (ImpResPAC Content Validation and Refinement).

Text has been added to the manuscript to clarify that the experts are invited to comment on the “Overview of ImpResPAC” and “ImpResPAC user instructions”. The sentence (line 230– 232) now reads:

“Experts will be asked to review the overview of ImpResPAC, ImpResPAC user instructions and ImpResPAC domain(s) and associated items for the domain(s) that they agree they are ‘experts’ in.”

On p.9, please also indicate whether in the second round of reviews, experts are going to review their “own” domains again or the whole tool.

We have amended as suggested (changes in bold). The sentence (line 238-243) now reads: All members of the ImpResPAC expert advisory group will be invited to review the refined version ImpResPAC and provide feedback on the acceptability, appropriateness and feasibility of ImpResPAC (all domains) via a follow-up survey. Experts will be given the option of providing feedback on the domains that they provided feedback on in Stage 1 (survey A) or if they wish, providing feedback on the entire tool.

P.10: The abbreviations “IF” and “D&I” are only used once and therefore don’t need to be introduced.

We have amended as suggested and removed IF and D&I (line 256-257).

On p.11, please indicate in the beginning of the assessment description of convergent validity whether the whole ImpResPAC or just certain dimensions will be correlated with the INSPECT scale.

We have amended as suggested. The sentence (line 297-299) now reads: “Convergent validity will be further examined by estimating the correlation between the global ImpResPAC scores with the global scores of INSPECT (14) as both scoring criteria rate the quality of proposed implementation science research. “

I found table 1 very helpful for a comparison of the dimensions of the two tools. However, I was surprised that there is no overlap between the INSPECT dimension “policy/funding environment” and the ImpResPAC dimension “determinants of implementation contextual factors”. Maybe the authors can provide a brief explanation for that.

We thank the reviewer for their feedback and acknowledge that given the Table 1 provides a high-level comparison of ImpResPAC and INSPECT (at a domain and element level respectively). It might appear surprising that there is no overlap between the INSPECT element “policy/funding environment” and the ImpResPAC dimension “determinants of implementation contextual factors”. The “Policy/funding environment” element of INSPECT does not overlap with ImpResPAC at a domain level. We believe this highlights the difference between the intended use of each tool, i.e., ImpResPAC is intended to appraise the conceptual and methodological quality of implementation research as opposed to fundability, that INSPECT intends to appraise. Although policy/funding environment is not a domain in ImpResPAC, that is not to say that policy and funding environment are not important to explore in implementation research, rather information on such would fall under the ‘determinants of implementation research: contextual factors’ domain. Table 1 has now been updated to provide more granular detail on the comparison of the two tool dimensions (INSPECT and ImpResPAC) and is included as supplementary material (additional file 2).

Discussion: On p.17, it is stated that the tool should help funders “differentiate high- and low-quality implementation research”. Since the tool contains several different domains to assess quality, I was wondering whether a more differentiated view on research quality would be more appropriate to describe the tool’s benefits, e.g., that it helps funders to “identify specific areas of improvement”.

We agree that in addition to differentiating high- and low-quality implementation research, ImpResPAC will also help funders to “identify specific areas of improvement”.

We have amended as suggested. The sentence (line 360-363) now reads:

“We envisage ImpResPAC primarily being used by funding bodies as a standardized and transparent method to differentiate high and low-quality implementation research and identify areas for improvement before funding decisions are made.”

Reviewer 2:

Thank you for submitting this protocol. It is clearly written and has potential to be a worthwhile extension of the ImpRes guide.

Thank you for the positive comments.

As lay reviewer, I was surprised that there is no public or patient involvement in the design of the study, or indeed intention to involve members of the public or patients as experts in any of the stages of the study. It would seem pertinent to involve them, especially given the ImpResPAC tool includes a PPI domain. PPI experts should represent a significant proportion of your target expert advisory group, as well as be included in the surveys.

We acknowledge the importance of public and patient involvement in the design of implementation research. Given that the study we report here is not funded, unfortunately we simply did not have the funds to involve patient and the public in the research. We acknowledge this is a limitation of our research and have included this as a limitation in the discussion section of the manuscript. We strongly recommend that any future ImpResPAC research, including further validation and utilisation, includes patient and public involvement.

We have amended the manuscript to include this limitation. The sentence (line 424-435) now reads: "We acknowledge the importance of public and patient involvement in the design of implementation research, but the study we report here is not funded and did not have the funds to involve patient and the public in the research. We strongly recommend that any future ImpResPAC research, including further validation and utilisation, includes patient and public involvement."

The proposal lacks detail on how the content of the existing INSPECT tool and the proposed ImpResPAC tool differ, other than at domain level. It is difficult, therefore, to see what are the assumed conceptual and methodological weaknesses INSPECT, which this proposal intends to review and enhance with ImpResPAC.

We acknowledge that the initial mapping of domains did not provide granular detail to adequately distinguish the two tools. Since stage 1 (mapping of ImpRes tool and guide to develop ImpResPAC domain items) is now complete, we are now able to provide a more detailed comparison of the ImpResPAC tool, at a domain item level and INSPECT tool, at an element item level. Please note, this comparison is an initial mapping and is subject to change after survey part A results have been analysed by the ImpResPAC team.

Supplementary material (additional file 2) provides a comparison of the ImpResPAC tool domain items and INSPECT element items.

We have added the following sentences to the manuscript (line 407-410) to direct readers to this comparison of the ImpResPAC tool and the INSPECT tool, at an item level:

"The initial mapping of the ImpRes tool and supplementation guide to develop the ImpResPAC tool (stage 1) and a detailed comparison of ImpResPAC tool domain items (initial mapping) and the INSPECT tool element items can be found in supplementary material (additional file 2)."

Without knowing the content of either tool, from my perspective, a head-to-head comparison of the "results" of using the two different tools on planned the case set of studies (in Stage 3), would be an important enhancement to this proposal in terms of its methodology.

We agree that providing a more detailed head-to-head comparison of ImpResPAC and INSPECT (presented in additional File 2) is useful for understanding the planned psychometric evaluation (detailed in stage 3 of our study) of ImpResPAC, specifically in relation to the assessment of the convergent validity of the ImpResPAC tool, which we plan to examine by correlating ImpResPAC tool domains and the INSPECT elements. In the manuscript this is described as follows (line 297-299):

"Convergent validity will be further examined by estimating the correlation between the global ImpResPAC dimension with the global scores of INSPECT (14) as both scoring criteria rate the quality of proposed implementation science research."

The proposal states that the application of the ImpResPAC tool in stage 3 would be done by the researchers. Is there opportunity to include the other possible end users in Stage 3, (categories as

per survey questions) namely those reviewing grant applications, practitioners, and educators. I would also suggest PPI reps here. Perhaps members of the expert advisory group could pilot / test the tool?

Although this is beyond the scope and funding of this study, we acknowledge and agree that inviting other end users to pilot and apply ImpResPAC (e.g., funders, educators and PPI members) would be a worthwhile in future study.

This limitation is acknowledged in the discussion of the manuscript (line 435-439):

“Secondly, we acknowledge that in order to truly test the value of ImpResPAC, it will be preferable to seek feedback from implementation research stakeholders who have had the opportunity to apply the tool in practice, but this is beyond the scope of this research. Future studies should evaluate the value of ImpResPAC with implementation research stakeholders that have applied the tool.”

The comparison of domains of INSPECT and the proposed ImpResPAC is helpful, but as mentioned before, without much detail (Table 1). I am guessing this is because it can be assumed that people will know the content of both INSPECT and the ImpRes guide. Nevertheless, the comparison seems a bit arbitrary, and may need just to be reviewed. Will this potentially change during Stage 1 of the study? In other words, should this be presented as an initial mapping, to be refined, by Stage 1 and then by survey part A?

We acknowledge that the initial mapping of ImpResPAC domains and INSPECT elements was a high-level overview. As stage 1 (mapping of ImpRes tool and guide to develop ImpResPAC domain items) is now complete, we are now able to provide a more detailed comparison of the ImpResPAC domain items and INSPECT element items. See supplementary material (additional file 2).

We have amended as suggested and the manuscript text has been updated to clearly state that the ImpResPAC tool presented is an ‘initial mapping’. The sentence (line 407-410) now reads:

“The initial mapping of the ImpRes tool and supplementation guide to develop the ImpResPAC tool (stage 1) and a detailed comparison of ImpResPAC tool domain items (initial mapping) and the INSPECT tool element items can be found in supplementary material (additional file 2)”

Also, the comparison in Table 1 does seem to suggest that INSPECT has several specific domains about context: care and quality gap; setting's readiness; policy environment. On the other hand the proposed ImpResPAC appears to roll these into one domain - contextual factors - albeit in the ImpRes guide this is about a wide range of contextual factors.

Stage 1 (mapping of ImpRes tool and guide to develop ImpResPAC domain items) is now complete, and highlights that INSPECT and ImpResPAC contain similar items (e.g., ‘The care or quality gap’) but map onto different domains (e.g., ‘The care or quality Gap’ is one of the ten INSPECT elements but falls under the ImpResPAC ‘Research Characteristics’ domain).

Furthermore, some items in INSPECT (e.g., ‘Team experience with setting, treatment, and implementation process’ and ‘Policy/funding environment; leverage of support for sustaining change’) do not overlap with any of the ImpResPAC domains items. We expect some difference in INSPECT and ImpResPAC items and the domains in which they fall. We believe these differences highlight the different focuses of is ImpResPAC (conceptual and methodological quality of implementation research) and INSPECT (fundability of implementation research).

We now have included an additional appendix showing a comparison of the ImpResPAC tool (initial version developed in stage 1) and the INSPECT tool.

The following text (line 407-410) has been added to introduce the initial mapping of the ImpResPAC tool (stage 1), presented as a comparison to the INSPECT tool:

“The initial mapping of the ImpRes tool and supplementation guide to develop the ImpResPAC tool (stage 1) and a detailed comparison of ImpResPAC tool domain items (initial mapping) and the INSPECT tool element items can be found in supplementary material (additional file 2)”

It is not clear when it is intended to run the study.

We have amended as suggested and have included date ranges in brackets after the sub-heading for stages 1-3 (line 190, 202-203 and 246-247).

Also, data ranges should be considered for the inclusion and exclusion criteria, or an explanation of why open dates are being used.

Two of the study authors (CS and LH), with expertise and experience in implementation and improvement science research, will independently appraise the conceptual and methodological quality of the 50 most recently published protocols going back in time from the date of the search and Implementation Science journal was launched in 2006.

What is the planned reach of the survey, in terms of range of stakeholders, countries, numbers ?

We have added the following sentence (line 221-223):

“We expect to identify 70 - 100 experts globally in the field of implementation science. We hope experts, both academics and practitioners, working in high-, middle- and low-income countries will participate.”

Reviewer: 3

This sequential mixed-methods study is designed to develop a quantitative tool for appraising the quality of implementation studies. I believe this tool could benefit the field.

Thank you for the positive comments.

Authors note that a lack of practical guidance likely contributes to poor quality implementation research. I'm wondering if there are examples or evidence that suggests that offering practical tools for study design improves research quality (if not in implementation, then in other fields). This would bolster the justification for this type of tool. For instance, there are many tools and resources for program evaluation – do we know if this actually improves the quality of evaluations? In implementation, we've seen many journals require that authors adhere to set reporting guidelines. Does this actually improve quality (at least of dissemination)?

This is a very valid point. Although there is evidence that tools to improve the quality of reporting improve reporting (e.g., the development of the Consolidated Standards of Reporting Trials (CONSORT) checklist, for the reporting of randomised controlled trials (RCTs), and Standards for Reporting Implementation Studies (StaRI) for implementation research), as far as we are aware, there is not the same level of evidence, that design tools, such as ImpRes, improves research quality or tools, like ImpResPAC, improves the quality of evaluation of implementation research.

As reported in the initial ImpRes development and evaluation manuscript (Hull et al, 2019), evaluation data indicated that experienced applied health researchers believe that ImpRes is likely to lead to better designed implementation studies. However, we acknowledge that more robust evidence is needed to assess whether the ImpRes tool and guide results into better designed implementation research.

We have updated the manuscript (line 117-120) to include the following sentence:

“Practical tools to improve the quality of reporting have been shown to improve research reporting (e.g., the development of the Consolidated Standards of Reporting Trials (CONSORT) checklist, for the reporting of randomised controlled trials (RCTs), (16) (17) (18).”

I also think it would be helpful to define what “high quality” implementation research is, from who's perspective, and whether there is consensus (or not) in the field. Given the broad range of studies I'd imagine that definitions vary widely across stakeholder groups. For instance, my funders might expect methodological rigor and somewhat controlled designs, while my community partners value applicability/utility. Does the inclusion of “experts” as defined potentially exclude perspectives that

could be useful? How might this limit the applicability of your tool? And does this reify dominant standards of research?

The aim of the ImpRes tool and guide was to define “high quality” implementation research and build consensus in the field to guide researchers and practitioners on incorporating core principles and concepts of high-quality implementation science and applied health implementation research into their implementation project. Building on this work, we aim to develop the ImpResPAC tool to identify quality indicators, to support funders in identifying high versus low quality implementation research. The ImpResPAC tool is based on the ImpRes tool and guide, which was developed through a consensus-building brainstorming process involving an international multidisciplinary expert panel, and identification of key methodological and conceptual literature containing design guidance and recommendations.

In development of the ImpRes tool and guide, the authors conducted a literature review, and then identified key principles and concepts through brainstorming. The perspective of those who engaged, could have limited the tool’s content, which could also limit the content of ImpResPAC. We acknowledge that how ‘high-quality’ research is defined, in the finalised version of the ImpResPAC tool, will be based on the perspectives of the expert advisory panel. We anticipate the expert advisory panel will include academics and practitioners working within the field, both in high, middle and low-income countries. We hope that the panel will also include academic and practitioner that represent patient, service users and public members and/or identify as such. We plan to collect this demographic data.

That said, we fully acknowledge that how we have defined ‘experts’ in this study could exclude the perspectives of other stakeholder groups that could be useful and how the tool might be valued by groups excluded in the initial development process. We now acknowledge this as a limitation in the discussion.

We have updated the manuscript (line 440-442) to include the following sentence:

“Thirdly, our definition of ‘experts’ (someone widely recognized as a reliable source of knowledge, technique, or skill whose judgment is accorded authority and status by the public or his or her peers) could exclude useful perspectives of stakeholder groups.”

The methods are staged thoughtfully. The description of Stage 1 suggests that psychometric testing will be conducted at this point of the study, but I think that actually happens in Stage 3. This could be clarified further. Additional detail about how the tool will be refined/developed would be useful.

To avoid confusion, we have amended as suggested (line 190). Psychometric evaluation is no longer mentioned in the section describing ‘Stage 1: ImpResPAC development’.

Finally, while I think practical tools for improving rigor could be extremely useful in a young and growing field, it might be worthwhile to consider the potential unintended consequences of such a tool. Does formalizing these quality standards restrict or disincentivize innovation? Disadvantage particular types of topics, forms of research, designs, etc? Prevent the introduction/application of different theoretical perspectives? Does it allow for evolution of the field and what we consider rigorous? If so, are these tradeoffs acceptable for the field?

We thank the reviewer for their thoughtful comments and prompting us to consider the potential unintended consequences of applying ImpResPAC. We acknowledge that there is always the possibility that practical tools developed to improve the quality of research could have negative unintended consequences, especially if used incorrectly. Although we do not foresee ImpResPAC favoring disincentivizing innovation within the field or particular topics, forms of research, or design, this is something that we, and other tool developers, must be conscious of and consider when designing and evaluating practical tools like ImpResPAC. Whilst beyond the initial development of the tool reported here, we recommend that future research evaluates whether applying ImpResPAC results in any unintended consequences, both positive and negative. If negative consequences of applying ImpResPAC are uncovered, solutions to minimize these must be developed.

VERSION 2 – REVIEW

REVIEWER	Marie-Therese Schultes University of Zurich, Institute for Implementation Science in Health Care
REVIEW RETURNED	19-Sep-2022

GENERAL COMMENTS	<p>Thank you for the opportunity to review the revised manuscript 'Development and Psychometric Evaluation of the Implementation Science Research Project Appraisal Criteria (ImpResPAC) tool: A Study Protocol'. In my opinion, the manuscript has clearly improved in clarity and comprehensiveness. I also very much appreciate the authors' detailed responses to each reviewer comment, which made it clear how the feedback has been incorporated in their revision.</p> <p>When reading the revised manuscript, I spotted some typos, e.g., on p.3, line 71 (double "quality"), p.4, line 94 (comma between references), p. 5, line 116 (missing parenthesis at the end of the sentence), p.6, line 135 (singular use of patient outcomes) and p.11, line 268 (neglectable "and"). Before final submission, I would recommend a thorough proofreading of the manuscript.</p> <p>Altogether, I am looking forward to see this interesting study protocol in print and to be able to use the ImpResPAC tool in the future.</p>
-------------------------	---

REVIEWER	Toto Gronlund NIHR, James Lind Alliance
REVIEW RETURNED	01-Sep-2022

GENERAL COMMENTS	Thank you for addressing the reviewers' comments so thoroughly.
-------------------------	---

Correction: 'Development and psychometric evaluation of the Implementation Science Research Project Appraisal Criteria (ImpResPAC) tool: a study protocol'

Sweetnam C, Goulding L, Davis RE, *et al.* Development and psychometric evaluation of the Implementation Science Research Project Appraisal Criteria (ImpResPAC) tool: a study protocol. *BMJ Open* 2022;12:e061209. doi: 10.1136/bmjopen-2022-061209.

Above mentioned article is revised since it was published as Funding statement is now updated as given below:

This research was supported by the National Institute for Health Research (NIHR) Applied Research Collaboration (ARC) South London at King's College Hospital NHS Foundation Trust; and by King's Improvement Science, which offers co-funding to the NIHR ARC South London and is funded by King's Health Partners (Guy's and St Thomas' NHS Foundation Trust, King's College Hospital NHS Foundation Trust, King's College London and South London and Maudsley NHS Foundation Trust), and Guy's and St Thomas' Charity (grant number: NIHR200152). NS' research is further supported by the SPIRES research programme (Antibiotic use across Surgical Pathways - Investigating, Redesigning and Evaluating Systems), funded by the Economic and Social Research Council. NS is further funded by the National Institute of Health Research (NIHR) Global Health Research Unit on Health System Strengthening in Sub-Saharan Africa, King's College London (GHRU 430 16/136/54) using UK aid from the UK Government to support global health research. The views expressed in this publication are those of the author(s) and not necessarily those of the NIHR, the charities, the ESRC or the Department of Health and Social Care.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

© Author(s) (or their employer(s)) 2024. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

BMJ Open 2024;14:e061209corr1. doi:10.1136/bmjopen-2022-061209corr1

