

## 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

<b>TITLE (PROVISIONAL)</b>	What are important consequences in children with non-specific spinal pain? A qualitative study of Danish 9-12-year olds
<b>AUTHORS</b>	Lauridsen, Henrik; Stolpe, Anna; Myburgh, Cornelius; Hestbæk, Lise

### VERSION 1 – REVIEW

<b>REVIEWER</b>	Tie Parma Yamato Universidade Cidade de São Paulo, Brazil
<b>REVIEW RETURNED</b>	08-Mar-2020

<b>GENERAL COMMENTS</b>	<p>Thank you for the opportunity to review for BMJ Open. The study covers an important issue on children and adolescents' pain, an area in need of more research and understanding. I do agree that we need to develop more questionnaires for the pediatric population and qualitative research is an essential first step on this. I have only minor comments in order to improve the manuscript.</p> <p>1) The authors interviewed children aged 9-12 years old. The explanation in the "participants" section is fine to me; however, this could be a limitation for developing a questionnaire for children and adolescents (8-18 years). This is interesting to be added to the discussion section.</p> <p>2) I'm curious on what is the definition you used for "troublesome spinal pain"?</p> <p>3) Data analysis: were a second reviewer checked the transcription and the codes? If all analyses were done by one reviewer only this could be a limitation for the qualitative approach and should be acknowledged in the discussion.</p> <p>4) I am a bit confused with the secondary objective "to compare these consequences with the content of common adult questionnaires". I honestly don't see the point of doing it as it is obvious it would be different, and the introduction section perfectly show this. The qualitative data is so rich that authors should consider exploring a bit more and highlight this more on the manuscript and give less focus to the secondary aim in the results and conclusion.</p> <p>5) I would like more information on why the authors excluded children with diagnosed musculoskeletal disorders since this could be very much linked with troublesome spinal pain that has an impact on children's life. In my opinion this is confusing. Also, if you recruited children with pain, some of them could have undiagnosed musculoskeletal disorders.</p>
-------------------------	--

<b>REVIEWER</b>	Dr Rebecca Rachael Lee University of Manchester, United Kingdom
<b>REVIEW RETURNED</b>	21-Apr-2020

<b>GENERAL COMMENTS</b>	<b>Comments for the authors</b>	
	<p>Thank you for the opportunity to review this manuscript describing a qualitative study exploring the appropriateness and relevance of common adult questionnaire content (specifically about the consequences of living with pain) in a paediatric population with spinal pain. The manuscript is based upon the analysis of interviews and focus group data from young participants who were recruited from schools in Denmark. The data were analysed following a thematic approach before additional focus groups were conducted. The authors come to the conclusion that questionnaires in this field that are designed for adults are not directly fit for purpose with children and young people and that new questionnaires with more relevant and appropriate content (which is identified from qualitative data in this paper) are necessary. The authors have chosen an interesting patient group as many studies of pain in children generalise patients across chronic pain conditions so this approach is useful for identifying issues that are pertinent to this particular condition. The article fits within the scope of the journal and has findings which may be of interest to healthcare professionals, researchers and even schools managing those with spinal pain (however, these advantages are not discussed), although the authors mainly focus on the advantages of these findings in new tool development. I have put together a list of specific comments addressing each section of the</p>	

	<p>paper. The concerns I have about the paper are predominantly: unclear structure in the methods (particularly around sampling strategy/participants and describing Study 2[focus groups] in a complimentary way to the description of Study 1 [interviews]), some problematic content of particular sentences throughout the paper and key information missing in the introduction (a stronger justification of why new measures are needed would be helpful) and methods (more detail around the process of the analysis).</p>	
	<p><b>Abstract specific comments</b></p>	<p>Background-</p> <p>No comments</p> <hr/> <p>Design and setting-</p> <ul style="list-style-type: none"> <li>I think the authors should refer to their study as using semi-structured interviews with no mention of 'case study' as this makes it sound like a one-off singular interview that was conducted with one child.</li> </ul> <hr/> <p>Participants-</p> <ul style="list-style-type: none"> <li>This should detail the sampling strategy and the selection, but I believe the number of participants included and age range is part of the results section (unless the requirements of this particular journal are different).</li> </ul> <hr/> <p>Methods-</p> <ul style="list-style-type: none"> <li>See comments below about use of the word 'conriformatory' for describing the focus groups.</li> </ul>

		<p>Results-</p> <p>No comments</p>
		<p>Conclusions-</p> <ul style="list-style-type: none"> <li>The authors repeat that there were five themes which is not necessary. The authors should discuss the implications of the work and future directions, not reiterate similar information to that covered in the results section.</li> </ul>
	<p><b>Introduction</b></p>	<p>The authors make a strong case for improving the outcomes of children and young people with chronic pain in general, considering the individual and societal burden of the condition. It would be useful if there was some specific data on the burden of spinal pain in particular or at least an acknowledgement of why there is no literature on this if it isn't possible to include here in the manuscript?</p>
		<p>"Despite the increasing interest"- for those who aren't familiar with this particular pain area, it would be useful if the authors could elaborate on what way interest has increased- are there new agendas being set to address this, new standards being introduced, more literature recently on the area?</p>
		<p>Line 29/30. Do authors mean constrains or constraints?</p>
		<p>Before discussing how adult questionnaires and constructs don't map onto children and young people's views, it would be useful if the authors provide an argument for why it is necessary that we are able to measure and assess perceptions around</p>

		consequences better than we currently do in this group? What will that information be useful for? A better justification for exploring this would be good.
		Line 50-51-Are these measures specific to assessing perceptions around consequences? Clarification on the nature of these measures would be useful.
		The authors state that there is an argument for more 'generalistic questionnaires'. I don't think this is a useful phrase, especially given that the authors are arguing for measures specifically about consequences. I'm confused as to what the exact aim and benefit of creating new questionnaires for this group is.
		The authors mention several questionnaires which refer to function and activity limitations. Is this because they anticipate that this might be one of the main important consequences for children and young people. They should discuss this if they think so.
		The justification for using qualitative methods appears warranted and beneficial.
	<b>Methods</b>	Il think the way the design section is written is quite misleading. The authors state that it is a qualitative case study (usually meaning one child) when there was more than one interview conducted. The authors also refer to 'focus group interviews'. Focus groups and interviews are two separate and very different methodologies, they should not be referred to as one and the same. I would also suggest the authors do not use the term

		'case study'.
		What do authors mean by 'the design focused on the child's cognitive level'? Do they mean that the interviews were conducted in a developmentally appropriate manner? If so, how did they ensure the questions were developmentally appropriate? Or did authors mean that during the interview, children and young people were asked questions around cognition, findings from were used as results? The meaning of this sentence is unclear.
		What do the authors mean by 'transformed'? Can they clarify this? Do they mean the themes were mapped onto existing classification systems? If so, they should say this more clearly. Transformed is ambiguous.
		I don't believe that focus groups should be used to 'confirm' anything. To confirm something would take a lot more quantitative investigation. I do however believe that focus groups can be useful for 'broadening' results and perspectives. The authors should reconsider their phrasing around the purpose of the focus groups.
		The authors should state their epistemological standpoint (constructivist) at the beginning of the design section if they are going to state it, but I don't think it is necessary as it shows through the methods that they have chosen for this work.
		Sampling strategy- it would be good to see some justification for why the authors recruited 9-12 year olds earlier in the strategy section. Also, authors argue that boys of 14 years have a growth spurt associated with spinal pain but then they do

		not include this age range in the sample. Can the authors find a justification that supports their age range chosen?
		The authors should re-think the phrase 'too affected by puberty'. This sounds like puberty is a troubling condition/illness in itself which is not the case.
		I don't think the phrase 'inside perspective' is appropriate. The authors asked children about their perceptions rather than proxies because proxies are often unreliable. Children reporting on their own perspectives does not provide an 'inside' look, but it provides arguably a more reliable account.
		'Interviews were conducted in each age band and well-being strata'. What were the age bands and the well-being strata that were chosen? 9-12 year olds already seems like quite a limited age band, did the authors deduce even smaller age ranges from this?
		New sentence should start after 'country', Line 10, Page 6.
		It is not very clear how authors recruited only those children who had spinal pain from the schools. Who identified those who did and did not have spinal pain from the result of the YSQ and kidscreen and therefore eligibility for the study? I also think this information should form part of the sampling strategy, not the participants section.
		The authors are more specific about the school included in the pilot compared to the actual study. Does revealing the more specific location of the school on the island affect the anonymity

		of the children involved?
		'To ensure all procedure were suitable for the study'- do authors mean they piloted everything including the recruitment strategy etc or just that they piloted the interview schedule? They should clarify this.
		What was the information material that was distributed to teachers, children and parents? What was the content/format of these and who created it? Do the authors mean information about the study or information about spinal pain? Similarly, what was the additional information that was provided to parents of children who were eligible to take part?
		The authors should define the ages of those in the 3 <sup>rd</sup> -6 <sup>th</sup> grades for readers from countries unfamiliar with the Danish schooling system.
		Was assent taken from children and young people? This should be discussed.
		What experience of qualitative methods did author ABS have? What was their background? It's important to describe the interviewers experience level in qualitative studies.
		'The face to face approach was chosen to meet the children's level of understanding'- I don't believe being face to face has any impact upon whether a child understands concepts better or not. Also whether face to face gets more co-operation from children and young people is debatable- research has started to suggest that young people talk more freely over the phone or other technologies where there is no direct eye



		<p>contact with the researcher. This whole sentence is problematic and based upon incorrect assumptions and weak justifications. The authors should rethink the advantages they cite for using face-to-face methodology and use literature to support their decision.</p>
		<p>I am unsure as to why the authors use the full body drawing and face pain scale revised (correct abbreviation for which is FPS-R, not rFPS). These are measures of pain intensity, which as far as I am aware is not a part of this studies aim. The interviews have been agued throughout the manuscript as being the method of exploring how children and young people express themselves- so it seems unjustified that then uni-dimensional pain intensity measures then appear in the methods? The authors are not interested in levels of pain so even if authors have used these tools to aid expression, they are being used to express the wrong thing. In Table 2, it says that participant had a score over 2, but in the inclusion criteria it says children had to score at least 3 on the FPS-R. The use of these measures and the purpose of including them in the study methodology requires a better description from the authors.</p>
		<p>Authors should provide a copy of the interview schedule within the manuscript or as supplementary material.</p>
		<p>Who transcribed the data? How were data transcribed, word for word, including the level of detail for a conversation analysis (pauses etc)?</p>

		<p>Were data anonymised or pseudoanonymised? Usually quite difficult (and sometimes unethical) to fully anonymise interview data for analytical purposes.</p>
		<p>Unusual that only one author generates themes, especially given that there is no discussion of these amongst the wider research team (or none that is mentioned). This should be discussed as a limitation of the study if there was only one researcher conducting the whole analysis, or at least reference to researcher bias in the discussion.</p>
		<p>Overall the methods section is not very well structured. The authors should consider using more appropriate subheadings and re-ordering parts of information that belong to those subheadings e.g. part of the sampling strategy is in the participants section, there is no materials/measures section, little information on the development of the interview schedule. The fact that interviews were semi-structured should come first in the design section, not in the data collection section. The structure also becomes complicated by the fact that authors have conducted 2 studies (interviews, then focus groups). Better structure needs to be in place to help the reader understand how the participant sampling strategies for both studies were different.</p>
		<p>Again, I don't think the focus groups 'confirm' anything, they just broaden the themes and methods.</p>
		<p>Do the authors mean 'break' rather than 'brake'?</p>

		<p>What was the wording given to children about sorting cards into relevant and irrelevant? This task sounds too difficult for that age range to understand as it is currently described in the paper.</p>
		<p>'probing for new codes to emerge'- this is problematic and suggests that the authors were performing a fishing exercise to find issues not highlighted in the interviews, artificially broadening their themes rather than allowing new themes to be identified by the participants themselves. Is this the case (not good practice if so!) or does the phrase just need rewording?</p>
		<p>Asking children and young people to help with theme categorisation seems a very complex task for the age range. Could authors provide the reader with examples of the wording so that we can be confident the children understood what they were being asked to do? I'm interested how the children came up with themes such as 'axial loading' which seems like a very sophisticated level of phrasing and understanding for 9-12 year olds.</p>
		<p>Who created the A3 drawing and words and what were the content? Was it stimulus created by children and young people themselves or something that the authors put together to represent what children and young people had said in the initial interviews?</p>
		<p>'consensus-based approach'- who gave the consensus, the researchers or children/young people or parents? This approach should be described in more detail, step-by-step, to allow someone to replicate it if</p>

		they wished so.
		There is no literature provided around thematic analysis- what approach to thematic analysis did the authors use, what where the steps followed, where are the references to thematic analysis literature they used to guide their analytical process?
		The authors have used the reporting checklist for qualitative studies but have not referenced that they have done this in the methods section. I am also not sure whether the document has been completed correctly, especially given that key information on the data analysis approach is missing. The authors should check over this again and make sure they have addressed each point thoroughly, as well as include a phrase about adhering to the checklist within the methods section.
	<b>Results</b>	Its unusual that a pilot results in no changes to the methodology. Especially given that patients/public were not involved in the design of the study. Could the authors elaborate on the findings of the pilot study?
		Why were only 36 children out of the 199 included? Was this to do with the randomised pick? Why did you stop data collection at that point? Would be good to recap your reader here.
		Not sure why a re-interview may have been necessary as an option as not discussed as part of the methods- could the authors clarify what would have constituted a re-interview?
		There is some overlap in themes e.g. about axial loading being a theme in itself and also

		a sub-theme of coping strategies. It would be good if there was a synthesis and narrative about overlapping subthemes.
		The themes all look plausible and there are some interesting consequences and impacts of living with spinal pain that are identified from participant accounts.
		Sixteen children in one focus group is a lot of participants. The authors should discuss how they managed these interactions and provide a rationale for why this didn't end up being two different focus groups, were discussions would have been arguably managed better.
		'It was possible to make all children partake in discussion'- the authors should tone down use of the word 'make' which suggests force or coercion.
		'Large uncertainty from children about fitting codes into themes'- do the authors think this might be because it was too complex of a task for the children and young people? Some reflection on some of these issues would be good.
		It would be useful to see supporting quotations and interview excerpts in the thematic analysis table as currently unsure what the particular purposes of Table 3 and 4 are. Can they be amalgamated?
	<b>Discussion</b>	The authors refer to a different aim in the beginning of their discussion compared to what is originally set out. In the discussion, authors refer to the paper identifying constructs of spinal pain important to children (appearing more generic), whereas the purpose

		<p>throughout has been about the consequences of living with spinal pain in particular. The authors should be more consistent in stating the same aims and purpose throughout the entire manuscript.</p>
		<p>The authors do not discuss 'think aloud' methodology in the methods but it is mentioned in the discussion. This is a specific qualitative methodology in itself. If this was used in interviews/focus groups, then the authors should discuss the approach in the methods section. Semi-structured interviews and focus groups are not 'thinking aloud'.</p>
		<p>I think it's important to highlight how not all pain has consequences for children and young people and I think this should have been a theme, even more so because these were children who did have some level of pain. I think there is some consequences of the difficulties of not understanding the etiology of pain which is also an interesting concept the authors touch on. They should address this interesting discussion in more detail.</p>
		<p>What do the authors mean by non-trivial back pain? The discussion is the first time this definition/criteria is introduced.</p>
		<p>There is no discussion about any implications of the findings. E.g. these themes are useful for questionnaire development, but can they be useful indicators for other types of research in this field? Do they suggest anything else should be further explored in this group? Are they helpful to anyone else other than the researchers of this paper who have a clear interest in</p>

		developing new tools?
		I think the conclusion should be stronger and not just be a re-cap of the themes identified. The authors haven't found that spinal pain is complex, but that children and young peoples' perceptions about the consequences of living with spinal pain are complex. This one-off qualitative study does not mean these themes should be the sole questions on new questionnaires, but it does pave the way for more exploratory work about other important issues (maybe not about consequences/impact) that should be included in questionnaires for this group.
	<b>My summary of review-</b> There are some interesting qualitative findings and themes in this paper however, the authors should revisit the methodology section in particular, so that the reader has more confidence in the results presented. The authors should reconsider how they justify the aims and benefits of answering this research question and include more literature to support certain study decisions. Some re-writing of problematic phrases is also necessary.	

<b>REVIEWER</b>	Dr Linda Birt Faculty of Medicine and Health Sciences University of East Anglia United Kingdom
<b>REVIEW RETURNED</b>	02-May-2020

<b>GENERAL COMMENTS</b>	Thank you for the opportunity to review this paper, which explores children's experience of spinal pain. It is innovative to see this qualitative data mapped to ICF. This study is the first step in developing a validated children's spinal pain questionnaire. Overall, the paper was easy to read and the tables clear. Below are some comments that I hope will increase clarity of procedures and decisions.
-------------------------	---

	<p>1. Sample criteria states 9-12 but later the article states member checking not undertaken as children aged under 12 year, then table 2 suggests some children aged 13 years – please clarify sample age.</p> <p>2. Was the children’s pain cluster around any sites? You mention using body maps but these results are not reported.</p> <p>3. Ethical points: did the children have the option of opting out of the screening questionnaire as there is no mention of consent at this stage? Were the researchers known to the children or the recruitment site?</p> <p>4. There could be stronger justification for research design i.e. why was method deemed appropriate?</p> <p>5. Use of focus groups as a form of validation was excellent. Did they have different interpretations and was there any additive data. The term focus groups interviews is used suggesting more than one I suggest stick with ‘focus groups discussion’ which you use a little later</p> <p>6. You state an expert was brought in for any disagreement in research team over ICF linkage. Please include examples of disagreement so the reader can more clearly understand this process of refinement</p> <p>7. Why were those adult questions pain questionnaires selected?</p> <p>8. While it is certainly true that pain narratives are influenced by sociocultural factors would it not be possible see if similar childhood accounts of pain are reported in other countries, otherwise the relevance to readers is rather limited.</p> <p>9. Is reference 9 available in English?</p> <p>10. Tables are clear , although I do not think figure 1 adds anything as it simply portrays ‘standard’ qualitative analysis . Of more interest and to increase credibility would be an explanation of where expert advice was sought on linking codes to ICF and what insights came from the focus groups.</p> <p>11. A few typos</p> <ul style="list-style-type: none"> <li>• Page 3 Strengths and limitations change ‘with’ to that in first bullet point</li> <li>• Page 8 line 48 brake should be break</li> <li>• Page 13 line 44 think about rephrasing ‘it was possible to make all children participate’ as sounds rather coercive at the moment .</li> </ul>
--	--

**VERSION 1 – AUTHOR RESPONSE**

Reviewer 1

<p>Thank you for the opportunity to review for BMJ Open. The study covers an important issue on children and adolescents’ pain, an area in need of more research and understanding. I do agree that we need to develop more questionnaires for the pediatric population and qualitative research is an essential first step on this. I have only minor comments in order to improve the manuscript.</p>	
<p>1) The authors interviewed children aged 9-12 years old. The explanation in the “participants” section is fine to me; however, this could be a limitation for developing a questionnaire for children and adolescents (8-18 years). This is interesting to be added to the discussion section.</p>	<p>We agree that the age range is narrow and a limiting factor. This has been added to the ‘Strength and limitations’ section.</p>



2) I'm curious on what is the definition you used for "troublesome spinal pain"?	Good point and thank you. We have removed the word 'troublesome' as we have defined spinal pain according to our inclusion criteria.
3) Data analysis: were a second reviewer checked the transcription and the codes? If all analyses were done by one reviewer only this could be a limitation for the qualitative approach and should be acknowledged in the discussion.	This is a good point which was also raised by the other reviewer. There were in fact two authors who checked the data analysis, and this has been added to the manuscript.
4) I am a bit confused with the secondary objective "to compare these consequences with the content of common adult questionnaires". I honestly don't see the point of doing it as it is obvious it would be different, and the introduction section perfectly show this. The qualitative data is so rich that authors should consider exploring a bit more and highlight this more on the manuscript and give less focus to the secondary aim in the results and conclusion.	We acknowledge the point of view of the reviewer. The reason for the comparison to adult questionnaires is that we don't know if it is same or not. At first glance, several of the issue's children talk about appear similar to adults (e.g. axial loading and physical function), however, digging one step deeper, the specific issues are very different. By making the comparison we make it absolutely clear that adult questionnaires cannot be used by children, and this is the main reason we wish to keep it in the paper.
5) I would like more information on why the authors excluded children with diagnosed musculoskeletal disorders since this could be very much linked with troublesome spinal pain that has an impact on children's life. In my opinion this is confusing. Also, if you recruited children with pain, some of them could have undiagnosed musculoskeletal disorders.	<p>Thank you for the comment. When we refer to 'diagnosed musculoskeletal (MSK) disorders', we refer to conditions like Muscular Dystrophy, Guillian-Barre Syndrome, Congenital Myopathy, Juvenile Osteoporosis and the like. These conditions may of course be linked to spinal pain, however, we were only interested non-specific spinal pain as this is by far the most prevalent condition. Furthermore, children with a MSK diagnosis are often medicalized which may have altered their view of back pain, and we were not interested in this. We have changed spinal pain to non-specific spinal pain throughout the text and added examples to the inclusion criterion to clarify this.</p> <p>Yes, it is correct that some of the children could have undiagnosed MSK disorders. However, we consider the likelihood of this to be small and unlikely to have interfered with our results.</p>

Reviewer 2

<p>Thank you for the opportunity to review this manuscript describing a qualitative study exploring the appropriateness and relevance of common adult questionnaire content (specifically about the consequences of living with pain) in a paediatric population with spinal pain. The manuscript is based upon the analysis of interviews and focus group data from young participants who were recruited from schools in Denmark. The data were analysed following a thematic approach before additional focus groups were conducted. The authors come to the conclusion that questionnaires in this field that are designed for adults are not directly fit for purpose with children and young people and that new questionnaires with</p>	
--	--

<p>more relevant and appropriate content (which is identified from qualitative data in this paper) are necessary. The authors have chosen an interesting patient group as many studies of pain in children generalise patients across chronic pain conditions so this approach is useful for identifying issues that are pertinent to this particular condition. The article fits within the scope of the journal and has findings which may be of interest to healthcare professionals, researchers and even schools managing those with spinal pain (however, these advantages are not discussed), although the authors mainly focus on the advantages of these findings in new tool development. I have put together a list of specific comments addressing each section of the paper. The concerns I have about the paper are predominantly: unclear structure in the methods (particularly around sampling strategy/participants and describing Study 2 [focus groups] in a complimentary way to the description of Study 1 [interviews]), some problematic content of particular sentences throughout the paper and key information missing in the introduction (a stronger justification of why new measures are needed would be helpful) and methods (more detail around the process of the analysis).</p>	
<p><b>Abstract</b></p>	
<p><b>Design and setting</b> I think the authors should refer to their study as using semi-structured interviews with no mention of 'case study' as this makes it sound like a one-off singular interview that was conducted with one child.</p>	<p>This is a good point. It has been changed throughout the manuscript.</p>
<p><b>Participants</b> This should detail the sampling strategy and the selection, but I believe the number of participants included and age range is part of the results section (unless the requirements of this particular journal are different).</p>	<p>We agree. The sampling strategy has now been included under 'Participants'.</p>
<p><b>Methods</b> See comments below about use of the word 'confirmatory' for describing the focus groups.</p>	<p>We agree and thank you. The term 'confirmatory' was used in lack of a better word covering what we did. We have now called it 'Focus group interviews' and elaborated on the purpose of them.</p>
<p><b>Conclusions</b> The authors repeat that there were five themes which is not necessary. The authors should discuss the implications of the work and future directions, not reiterate similar information to that covered in the results section.</p>	<p>The conclusion has been changed to focus more on the differences between children and adults and what implications this has to future work when developing a questionnaire.</p>
<p><b>Introduction</b></p>	
<p>The authors make a strong case for improving the outcomes of children and young people with chronic pain in general, considering the individual and societal burden of the condition. It would be useful if there was some specific data on the burden of spinal pain in particular or at least an acknowledgement of why there is no literature on this if it isn't possible to include here in the manuscript?</p>	<p>We agree and have added a comment on this in the introduction from the Global Burden of Disease study.</p>
<p>"Despite the increasing interest"- for those who aren't familiar with this particular pain area, it would be</p>	<p>Good idea. We have added focus areas of recent systematic reviews with references to</p>

useful if the authors could elaborate on what way interest has increased - are there new agendas being set to address this, new standards being introduced, more literature recently on the area?	address this.
Line 29/30. Do authors mean constrains or constraints?	Well spotted. This is a typo and has been corrected. It should read constraints.
Before discussing how adult questionnaires and constructs don't map onto children and young people's views, it would be useful if the authors provide an argument for why it is necessary that we are able to measure and assess perceptions around consequences better than we currently do in this group? What will that information be useful for? A better justification for exploring this would be good.	We agree and have added a sentence about this in the introduction: 'Such questionnaires are important to measure outcomes in clinical trials as well as monitoring patient progress in clinical practice.'
Line 50-51-Are these measures specific to assessing perceptions around consequences? Clarification on the nature of these measures would be useful.	No, they are paediatric-specific pain questionnaires where pain is measured among several other dimensions. We have added the names of the questionnaires and what they measure.
The authors state that there is an argument for more 'generalistic questionnaires'. I don't think this is a useful phrase, especially given that the authors are arguing for measures specifically about consequences. I'm confused as to what the exact aim and benefit of creating new questionnaires for this group is.	Thank you for your comment. The 'generalistic' refers to the population, not what the questionnaire measures. The Paediatric Back Outcome Measure is intended for use in athletes, and we wanted to investigate a broad population of children with spinal pain. The sentence has been clarified to avoid misunderstanding.
The authors mention several questionnaires which refer to function and activity limitations. Is this because they anticipate that this might be one of the main important consequences for children and young people. They should discuss this if they think so.	We have briefly described the types of existing questionnaires developed specifically to children. These are mainly focusing on general pain or pain/symptoms/function related to specific diseases. One is specifically related to back pain in children. The reason for this is to show that the area of consequences of spinal pain in children/adolescents has not been investigated as the focus has been elsewhere, and this is the reason we are doing it.
The justification for using qualitative methods appears warranted and beneficial.	Thank you.
<b>Methods</b>	
Il think the way the design section is written is quite misleading. The authors state that it is a qualitative case study (usually meaning one child) when there was more than one interview conducted. The authors also refer to 'focus group interviews'. Focus groups and interviews are two separate and very different methodologies, they should not be referred to as one and the same. I would also suggest the authors do not use the term 'case study'.	This is a good suggestion and we agree. We have changed the methods into two separate studies: The interview study and the focus group study.
What do authors mean by 'the design focused on the child's cognitive level'? Do they mean that the interviews were conducted in a developmentally appropriate manner? If so, how did thy ensure the questions were developmentally appropriate? Or did authors mean that during the interview, children and young people were asked questions around cognition, findings from were used as results? The	We agree that this is unclear. We have removed the sentence as it is already described more clearly in the section on 'Data collection'. Here it reads: 'The face-to-face approach was chosen to meet the children's level of understanding and to get the best cooperation with the children.'

meaning of this sentence is unclear.	
What do the authors mean by 'transformed'? Can they clarify this? Do they mean the themes were mapped onto existing classification systems? If so, they should say this more clearly. Transformed is ambiguous.	You are correct. 'Transformed' is unclear and we have changed the wording of the sentence: 'Finally, results were mapped onto the International Classification of Function, Disability and Health...'
I don't believe that focus groups should be used to 'confirm' anything. To confirm something would take a lot more quantitative investigation. I do however believe that focus groups can be useful for 'broadening' results and perspectives. The authors should reconsider their phrasing around the purpose of the focus groups.	We agree. This has been addressed throughout the manuscript as stated above.
The authors should state their epistemological standpoint (constructivist) at the beginning of the design section if they are going to state it, but I don't think it is necessary as it shows through the methods that they have chosen for this work.	This was added as it is a requirement in the 'reporting checklist for qualitative studies'. We have therefore decided to leave it in the manuscript, however, we will ask for an editorial decision on this.
Sampling strategy- it would be good to see some justification for why the authors recruited 9-12 year olds earlier in the strategy section. Also, authors argue that boys of 14 years have a growth spurt associated with spinal pain but then they do not include this age range in the sample. Can the authors find a justification that supports their age range chosen?	This makes sense. We have moved the description of the justification of the age limits to the 'Sampling strategy' section.  We have deliberately chosen not to include children older than 12 years as this will avoid all the changes seen in puberty where the incidence of spinal pain increases. The included sample range from 9-12 years as described in Table 2. We did have children aged 13, but they were not included.
The authors should re-think the phrase 'too affected by puberty'. This sounds like puberty is a troubling condition/illness in itself which is not the case.	Thanks. This has been changed.
I don't think the phrase 'inside perspective' is appropriate. The authors asked children about their perceptions rather than proxies because proxies are often unreliable. Children reporting on their own perspectives does not provide an 'inside' look, but it provides arguably a more reliable account.	Agree. We have changed the sentence and avoided the concept of 'inside perspective'.
'Interviews were conducted in each age band and well-being strata'. What were the age bands and the well-being strata that were chosen? 9-12 year olds already seems like quite a limited age band, did the authors deduce even smaller age ranges from this?	We agree that the terminology 'age band' is confusing as 9-12 years of age is already narrow. This has been changed to 'age'. The children filled out the Kid-Screen, and we used the cut-off values from the published manual to stratify into low, normal and high levels of well-being. This is described under 'Participants'.
New sentence should start after 'country', Line 10, Page 6.	This has been changed.
It is not very clear how authors recruited only those children who had spinal pain from the schools. Who identified those who did and did not have spinal pain from the result of the YSQ and kidscreen and therefore eligibility for the study? I also think this information should form part of the sampling strategy, not the participants section.	The selection criteria for children with spinal pain are listed in Table 1. However, we have changed the wording to clarify this and moved the paragraphs to the 'Sampling strategy' section.
The authors are more specific about the school included in the pilot compared to the actual study. Does revealing the more specific location of the	Point taken. Anonymity was absolutely paramount to get the protocol approved by the Danish Data Protection Agency. We have

school on the island affect the anonymity of the children involved?	changed the description of the pilot procedures to the same as the main study, and we hope this is acceptable.
'To ensure all procedure were suitable for the study'- do authors mean they piloted everything including the recruitment strategy etc or just that they piloted the interview schedule? They should clarify this.	The pilot procedures tested all the procedures, and this included the recruitment strategy. We have changed the wording of the sentence to accommodate this.
What was the information material that was distributed to teachers, children and parents? What was the content/format of these and who created it? Do the authors mean information about the study or information about spinal pain? Similarly, what was the additional information that was provided to parents of children who were eligible to take part?	The general information material consisted of a layman description of the project and the selection procedures that would be implemented. The additional information provided a more information about the objectives and procedures for the interview the selected children would go through. All material was written by the author group and strictly comply with the GDPR. We have not included this material as it is in Danish. Instead we have clarified the description in the manuscript.
The authors should define the ages of those in the 3rd-6th grades for readers from countries unfamiliar with the Danish schooling system.	Thank you. In 'Data collection' under the 'Interview study' we have added a sentence explaining this.
Was assent taken from children and young people? This should be discussed.	We did not collect informed consent from the children as they are under 15. Instead we collected informed consent from the parents to the children, as prescribed by Danish law, and this is described under 'Data collection'.
What experience of qualitative methods did author ABS have? What was their background? It's important to describe the interviewers experience level in qualitative studies.	ABS was a novice qualitative researcher, but was supported throughout by CM, who has been conducting qualitative research for fifteen years. We have added that CM was part of the coding and interpretation process.
'The face to face approach was chosen to meet the children's level of understanding'- I don't believe being face to face has any impact upon whether a child understands concepts better or not. Also whether face to face gets more co-operation from children and young people is debatable- research has started to suggest that young people talk more freely over the phone or other technologies where there is no direct eye contact with the researcher. This whole sentence is problematic and based upon incorrect assumptions and weak justifications. The authors should rethink the advantages they cite for using face-toface methodology and use literature to support their decision.	Thank you. We have changed our rationale for choosing the face-to-face approach to the following: "The face-to-face approach was chosen to allow co-creation of meaning by reconstructing perceptions of experiences and events related to spinal pain [Dicicco-Bloom et al.]."
I am unsure as to why the authors use the full body drawing and face pain scale revised (correct abbreviation for which is FPS-R, not rFPS). These are measures of pain intensity, which as far as I am aware is not a part of this studies aim. The interviews have been agued throughout the manuscript as being the method of exploring how children and young people express themselves- so it seems unjustified that then uni-dimensional pain intensity measures then appear in the methods? The authors are not interested in levels of pain so even if authors have used these tools to aid expression, they	This was not how the two instruments were used. It is correct that FPS-R is a measure of pain intensity, however the body drawing is not. We used the body drawing to locate the painful area to make sure that we were talking about spinal pain. The FPS-R was only used to gage the amount of pain the child experienced to confirm the inclusion criteria. During the rest of the interview, the children were completely free to express themselves. We agree that this is unclear in the manuscript and have changed it

are being used to express the wrong thing. In Table 2, it says that participant had a score over 2, but in the inclusion criteria it says children had to score at least 3 on the FPS-R. The use of these measures and the purpose of including them in the study methodology requires a better description from the authors.	according to the description above.  Regarding the scores, we have changed the wording to rectify and avoid confusion.  The rationale for this inclusion criteria has now been described under the sampling strategy.
Authors should provide a copy of the interview schedule within the manuscript or as supplementary material.	We are happy to provide the interview schedule as an appendix, however, it will be in Danish. This is the reason we have not included it as we considered it irrelevant for most readers. Instead we have included a section called 'Semi-structured interview' where we have elaborated on the interview guide.
Who transcribed the data? How were data transcribed, word for word, including the level of detail for a conversation analysis (pauses etc)?	The interviews were recorded and transcribed verbatim by the co-author ABS. The transcripts included non-verbal communication cues.
Were data anonymised or pseudoanonymised? Usually quite difficult (and sometimes unethical) to fully anonymise interview data for analytical purposes.	The data were pseudoanonymised but stored on a secure server at the university.
Unusual that only one author generates themes, especially given that there is no discussion of these amongst the wider research team (or none that is mentioned). This should be discussed as a limitation of the study if there was only one researcher conducting the whole analysis, or at least reference to researcher bias in the discussion.	Good point. We chose to simplify the description of this part of the data analysis to shorten the manuscript. The coding was in fact carried out by two of the authors (ABS and CM) with ABS as the main coder. The themes were developed among all the authors with ABS taking the lead. We have expanded the description of this and hope it is acceptable.
Overall the methods section is not very well structured. The authors should consider using more appropriate subheadings and re-ordering parts of information that belong to those subheadings e.g. part of the sampling strategy is in the participants section, there is no materials/measures section, little information on the development of the interview schedule. The fact that interviews were semi-structured should come first in the design section, not in the data collection section. The structure also becomes complicated by the fact that authors have conducted 2 studies (interviews, then focus groups). Better structure needs to be in place to help the reader understand how the participant sampling strategies for both studies were different.	We have changed the structure of the methods section as suggested by the reviewer. Firstly, we have divided it into two separate studies – the interview study and the focus group study. Secondly, we have added new headings and changed existing headings, and thirdly, we have moved the descriptions of the method to the relevant headings. We are happy with the suggestions and believe the structure has improved considerably. We hope the reviewer agrees.
Again, I don't think the focus groups 'confirm' anything, they just broaden the themes and methods.	We agree. It has been changed.
Do the authors mean 'break' rather than 'brake'?	This is a typo and has been corrected to break.
What was the wording given to children about sorting cards into relevant and irrelevant? This task sounds too difficult for that age range to understand as it is currently described in the paper.	The words 'relevant' or 'irrelevant' are only used in the manuscript to describe what we did, as we agree that children of this age would not understand them. We carefully explained to the children (with examples) what they should do: Think of your own back pain and then take one card at a time and see if it applies to you. We have added some of

	this in the text for clarification.
'probing for new codes to emerge'- this is problematic and suggests that the authors were performing a fishing exercise to find issues not highlighted in the interviews, artificially broadening their themes rather than allowing new themes to be identified by the participants themselves. Is this the case (not good practice if so!) or does the phrase just need rewording?	We agree that this is unclear. We have used the wrong wording. This has been changed in the manuscript.
Asking children and young people to help with theme categorisation seems a very complex task for the age range. Could authors provide the reader with examples of the wording so that we can be confident the children understood what they were being asked to do? I'm interested how the children came up with themes such as 'axial loading' which seems like a very sophisticated level of phrasing and understanding for 9-12 year olds.	The themes used during the focus group study were developed by the research group. These were too sophisticated for the children to understand. Therefore, we developed A3 posters with theme words and a visual drawing representing the theme. In addition, we explained each theme and gave examples of how codes could be categorized into a certain theme. We did not ask the children to develop themes but only to place the cards with codes on the theme which they found appropriate. However, we paid attention to any new codes which emerged. This has been clarified in the text.
Who created the A3 drawing and words and what were the content? Was it stimulus created by children and young people themselves or something that the authors put together to represent what children and young people had said in the initial interviews?	The author team generated the A3 posters. The words for each theme were identified from the codes in the interview study, and from this the author team developed one-two visual drawings representing the theme. We have clarified this in a new 'Materials' section in the focus group study.
'consensus-based approach'- who gave the consensus, the researchers or children/young people or parents? This approach should be described in more detail, step-by-step, to allow someone to replicate it if they wished so.	Thank you. We have elaborated on this in the text.
There is no literature provided around thematic analysis- what approach to thematic analysis did the authors use, what were the steps followed, where are the references to thematic analysis literature they used to guide their analytical process?	We agree that this needs to be added. The thematic approach used is the one outlined by Braun & Clarke and involves five steps: <ol style="list-style-type: none"> <li>1. Data Familiarisation</li> <li>2. Generating initial codes</li> <li>3. Searching for themes</li> <li>4. Reviewing themes</li> <li>5. Define and name themes</li> </ol> <p>This has been added to the 'Data analysis' section of the 'Interview study'.</p> <p>Braun, V., &amp; Clarke, V. (2006). Using thematic analysis in psychology. <i>Qualitative research in psychology</i>, 3(2), 77-101.</p>
The authors have used the reporting checklist for qualitative studies but have not referenced that they have done this in the methods section. I am also not sure whether the document has been completed	We have reviewed the reporting checklist with respect to all the changes made to the manuscript and ensured that all the points are addressed. We have also mentioned the use

correctly, especially given that key information on the data analysis approach is missing. The authors should check over this again and make sure they have addressed each point thoroughly, as well as include a phrase about adhering to the checklist within the methods section.	of it in methods section.
<b>Results</b>	
Its unusual that a pilot results in no changes to the methodology. Especially given that patients/public were not involved in the design of the study. Could the authors elaborate on the findings of the pilot study?	Yes, we agree that it is a little unusual, however, careful planning made the pilot test very successful. We did make minor adjustments, but to save space we did not include this in the first version. We have now elaborated more on this.
Why were only 36 children out of the 199 included? Was this to do with the randomised pick? Why did you stop data collection at that point? Would be good to recap your reader here.	Point taken. We have added clarification to the paragraph on 'Participants' in the results section.
Not sure why a re-interview may have been necessary as an option as not discussed as part of the methods- could the authors clarify what would have constituted a re-interview?	No, you are absolutely right. This has been removed.
There is some overlap in themes e.g. about axial loading being a theme in itself and also a sub-theme of coping strategies. It would be good if there was a synthesis and narrative about overlapping subthemes.	Yes, this is true. We have made a short comment on axial loading also being a subtheme of coping strategies in the text.
The themes all look plausible and there are some interesting consequences and impacts of living with spinal pain that are identified from participant accounts.	Thank you.
Sixteen children in one focus group is a lot of participants. The authors should discuss how they managed these interactions and provide a rationale for why this didn't end up being two different focus groups, were discussions would have been arguably managed better.	This is well spotted. In fact, we did divide the children into two focus groups for the first part sorting the cards into two piles and the subsequent group-discussion. After this we collapsed both focus groups into one to see if further information could be obtained by the groups inspiring each other. We have changed the description in the manuscript to match this.
'It was possible to make all children partake in discussion'- the authors should tone down use of the word 'make' which suggests force or coercion.	This has been changed.
'Large uncertainty from children about fitting codes into themes'- do the authors think this might be because it was too complex of a task for the children and young people? Some reflection on some of these issues would be good.	Thank you. We have elaborated on this in the text.
It would be useful to see supporting quotations and interview excerpts in the thematic analysis table as currently unsure what the particular purposes of Table 3 and 4 are. Can they be amalgamated?	Since the ICF classifications and the themes do not exactly overlap, we find it difficult to combine Tables 3 and 4. We also find it redundant to include quotes in Table 4 as they are included in Table 3 for all codes.
<b>Discussion</b>	
The authors refer to a different aim in the beginning of their discussion compared to what is originally set out. In the discussion, authors refer to the paper identifying constructs of spinal pain important to children (appearing more generic), whereas the purpose throughout has been about the	We agree and have corrected the beginning of the introduction.



consequences of living with spinal pain in particular. The authors should be more consistent in stating the same aims and purpose throughout the entire manuscript.	
The authors do not discuss 'think aloud' methodology in the methods but it is mentioned in the discussion. This is a specific qualitative methodology in itself. If this was used in interviews/focus groups, then the authors should discuss the approach in the methods section. Semi-structured interviews and focus groups are not 'thinking aloud'.	This is a writing error. We did not use the 'think aloud' methodology and this was purely an observation by the authors. It has been removed from the discussion.
I think it's important to highlight how not all pain has consequences for children and young people and I think this should have been a theme, even more so because these were children who did have some level of pain. I think there is some consequences of the difficulties of not understanding the etiology of pain which is also an interesting concept the authors touch on. They should address this interesting discussion in more detail.	We agree that these are important issues, and we have included them in 'Salient issues' in the results section. The reason for not including 'no limitations' as a theme is that it does not describe consequences of pain. We discuss possible reasons for this in the discussion and to highlight the issue further, we have added it to the conclusion.
What do the authors mean by non-trivial back pain? The discussion is the first time this definition/criteria is introduced.	Point taken. Non-trivial pain is now defined in the sampling strategy and we refer to the inclusion criteria of the study. We hope this clarifies the confusion.
There is no discussion about any implications of the findings. E.g. these themes are useful for questionnaire development, but can they be useful indicators for other types of research in this field? Do they suggest anything else should be further explored in this group? Are they helpful to anyone else other than the researchers of this paper who have a clear interest in developing new tools?	This is a good point and we have added a sentence to the first paragraph of the discussion elaborating on this.
I think the conclusion should be stronger and not just be a re-cap of the themes identified. The authors haven't found that spinal pain is complex, but that children and young peoples' perceptions about the consequences of living with spinal pain are complex. This one-off qualitative study does not mean these themes should be the sole questions on new questionnaires, but it does pave the way for more exploratory work about other important issues (maybe not about consequences/impact) that should be included in questionnaires for this group.	We agree and have changed the conclusion. We have removed the five themes and highlighted that the codes differ compared to adult questionnaires. In addition, we now recommend that the identified themes and codes as a starting point for the development of a new questionnaire.
<b>My summary of review-</b> There are some interesting qualitative findings and themes in this paper however, the authors should revisit the methodology section in particular, so that the reader has more confidence in the results presented. The authors should reconsider how they justify the aims and benefits of answering this research question and include more literature to support certain study decisions. Some re-writing of problematic phrases is also necessary.	Thank you. We have followed the advice of the reviewer to the best of our ability.

Reviewer 3

Thank you for the opportunity to review this paper, which explores children's experience of	
---	--

<p>spinal pain. It is innovative to see this qualitative data mapped to ICF. This study is the first step in developing a validated children's spinal pain questionnaire. Overall, the paper was easy to read and the tables clear. Below are some comments that I hope will increase clarity of procedures and decisions.</p>	
<p>1. Sample criteria states 9-12 but later the article states member checking not undertaken as children aged under 12 year, then table 2 suggests some children aged 13 years – please clarify sample age.</p>	<p>Thank you and well spotted. We have changed the text making sure that it is consistent throughout the paper. Table 2 includes children aged 13 as the 6. grade includes this age group, and we would show incomplete data if we did not include them. However, we did not include any of them which is also clear from Table 2.</p>
<p>2. Was the children's pain cluster around any sites? You mention using body maps but these results are not reported.</p>	<p>We used pain drawings to ensure that the painful site was in fact the spine and not any other area (the children pointed to the painful site on the drawing). We did not record if they had pain in other areas, however, the main painful site had to be the spine. This has been clarified in the revised manuscript.</p>
<p>3. Ethical points: did the children have the option of opting out of the screening questionnaire as there is no mention of consent at this stage? Were the researchers known to the children or the recruitment site?</p>	<p>Good point. Yes, the children could opt out at any time if they so wished. This was part of the information to the school and parents and follows the Danish Data Protection Agency's rules and regulations. The following sentence has been added to the 'Sampling strategy' section: 'The children were able to opt out at any time during the selection procedure and in the interview.' One researcher (ABS) was familiar with the school but none of the children were known to any of the researchers.</p>
<p>4. There could be stronger justification for research design i.e. why was method deemed appropriate?</p>	<p>Yes, we agree. We have divided the study into two parts on recommendation of one of the other reviewers: The interview study and the focus group study. In the 'Design' section we have justified the chosen methods for these two 'substudies', and we hope this is satisfactory.</p>
<p>5. Use of focus groups as a form of validation was excellent. Did they have different interpretations and was there any additive data. The term focus groups interviews is used suggesting more than one I suggest stick with 'focus groups discussion' which you use a little later</p>	<p>Thank you. We used the focus group study as a form of validation and elaboration of the findings from the interview study. Three of the authors participated in the focus group study, and we all agreed that the main codes and themes were confirmed, but that the children elaborated on them using slightly different words to describe what they felt. Regarding the number of focus groups, there were in fact two. We divided the children into two focus groups for the first part sorting the cards into two piles and the subsequent group-discussion. After this we collapsed both focus groups into one to see if further information could be obtained by the groups inspiring each other. We have changed the description in the manuscript to match this.</p>
<p>6. You state an expert was brought in for any disagreement in research team over ICF linkage.</p>	<p>Yes, our description needs to be clearer, and this has also been mentioned by another reviewer.</p>

Please include examples of disagreement so the reader can more clearly understand this process of refinement	Instead of giving examples, we have explicitly clarified how we did it, so the procedure can be reproduced by other researchers. We refer the reviewer to the section 'Code mapping to the ICF and comparison to adult questionnaires'.
7. Why were those adult questions pain questionnaires selected?	Questionnaires for spinal pain only exists for the low back and the neck. The ones chosen for these regions were selected as 1) they are the most commonly used, and 2) there were studies categorizing their content according to the ICF. This has been clarified in the manuscript.
8. While it is certainly true that pain narratives are influenced by sociocultural factors would it not be possible see if similar childhood accounts of pain are reported in other countries, otherwise the relevance to readers is rather limited.	This is a good point. So far, we are not aware of any other studies which report on pain narratives for this age group. Instead we have added the following sentence:  'Therefore, we recommend that studies on qualitative pain narratives in this age group are carried out in other cultures and the results compared to ours.'
9. Is reference 9 available in English?	Unfortunately not. This is a Danish book outlining research on children as equal partners. The translated title is:  'Child perspectives: Children as equal partners in social and educational work'  We have changed the Danish title to the English and included that it is a Danish book in the reference.
10. Tables are clear, although I do not think figure 1 adds anything as it simply portrays 'standard' qualitative analysis. Of more interest and to increase credibility would be an explanation of where expert advice was sought on linking codes to ICF and what insights came from the focus groups.	We see the point made by the reviewer. However, to give the unenlightened reader a chance to understand our procedure, we have chosen to keep Figure 1 as is. However, we have added extra explanation about the expert used to resolve issues with the linking codes and hope this satisfies the reviewer.
11. A few typos • Page 3 Strengths and limitations change 'with' to that in first bullet point • Page 8 line 48 brake should be break • Page 13 line 44 think about rephrasing 'it was possible to make all children participate' as sounds rather coercive at the moment .	Thank you, well spotted. Everything has been changed.

### VERSION 2 – REVIEW

<b>REVIEWER</b>	Dr Linda Birt University of East Anglia England
<b>REVIEW RETURNED</b>	25-Jun-2020
<b>GENERAL COMMENTS</b>	Thank you for your full responses to my and other reviewers comments. You have addressed all points and the methodology within this study is now much more transparent.