

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)	Ultrasound measurement of traumatic scar and skin thickness: A scoping review of evidence across the translational pipeline of research-to-practice
AUTHORS	Meikle, Brandon; Simons, Megan; Mahoney, Tamsin; Reddan, Tristan; Dai, Bryan; Kimble, Roy; Tyack, Zephaniae

VERSION 1 – REVIEW

REVIEWER	Malic, Claudia University of Ottawa, DEpt of Surgery
REVIEW RETURNED	27-Sep-2023

GENERAL COMMENTS	<p>This is a scoping review to evaluate the use of USS in measuring the scar thickness. the description of the methodology is clear and also the results are explained.</p> <p>The introduction is a bit long and although i appreciate for the description of the USS and the problem, it should be more concised.</p> <p>Another suggestion will be to use the word manuscript instead of records.</p> <p>At the beginning of the results it was mentioned about the number of patients in all the studies but there were no further results related to this. I would prefer to see this elaborated a bit more in relations to the headings that were in the results. Although it is important how many publications addressed an item, it is also important the size of the cohorts. the search algorithm and the PRISMA should be placed as supplementary material. A table with the size of the cohorts for each subheading will be useful.</p>
-------------------------	--

REVIEWER	Yao, Min Shanghai Jiao Tong University School of Medicine
REVIEW RETURNED	01-Oct-2023

GENERAL COMMENTS	I recommend to accept the paper for publication.
-------------------------	--

REVIEWER	Kennedy, Donna Imperial College Healthcare NHS Trust, Therapy Department
REVIEW RETURNED	22-Oct-2023

GENERAL COMMENTS	Thank you for the opportunity to review this interesting and timely manuscript. Herein the authors perform a scoping review to identify what is known and unknown about measurement of traumatic cutaneous scar thickness using ultrasound. The broad scope of the review results in a mass of references and associated data that are difficult for the reader to interpret as presented.
-------------------------	--

	<p>Please find observations and relevant comments to support editing this important manuscript below.</p> <p>1. The search strategy, per published protocol, identified peer reviewed literature as an inclusion criteria. However, the authors have included abstracts, which account for 30% of the included references. Unless demonstrated otherwise, the assumption would be that many abstracts result from conference posters and presentations; generally not peer reviewed work, therefore should not have been included. To this point, in your abstract conclusion you note there is poor methodological reporting of included manuscripts. But this seems possibly an unfair assertion, if 30% of the body of work are abstracts with requisite abbreviated methods. This point is highlighted by the frequency of denoting 'not reported' in table 2. Perhaps consider if you can justify including the abstracts and if they add to or detract from the reviews take away messages. Additionally, the scope of the review is evidence across the translational pipeline, but studies in humans was in inclusion criteria. Therefore no phase one basic or pre-clinical evidence was included. Perhaps clarify the scope of the paper relative to the search strategy? Lastly, non-traumatic scarring, such as acne scarring, was a review exclusion. However, in table 1, acne scarring is used to define the "mixed scarring" category. If excluded conditions are included in "mixed scarring" cases, does this have the potential to overinflate the total cases or patients involved in the review?</p> <p>2. The data extraction fields reported in supplementary table 1 are comprehensive, and it would be helpful to have more of this data included in the manuscript to improve interpretability. For example, in table 2, perhaps list included studies alphabetically but by study phase. Additionally, include the country where the work was completed, funding sources (i.e. commercial or non-commercial) and where scar aetiology is surgical, reporting the surgical cohort or condition. This will enable the reader to better appreciate gaps and challenges to progressing US to fruition as a clinical measurement tool and may promote collaboration and cross-fertilization, as appears to be the overarching remit of the review.</p> <p>3. Abstract, line 33. methods are reported in the literature. The abstract results and conclusion appear to omit important findings, for example reliability data. The conclusion centres on the authors' recommendations around standardising measurement, whereas my take on this is that the authors have extracted information to support standardisation and curated supplementary table 4.</p> <p>4. Introduction line 78 - How is a "major" skin injury defined?</p> <p>line 98 - While it is interesting, the relevance of the skills required to perform and interpret ultrasound as related to the review aims is not clear.</p> <p>line 124 - this reads as though US has successfully been implemented to support personalised medicine. Perhaps elaborate on what treatment protocols you are referring to? Does this apply to all scars and all interventions?</p> <p>line 132 - methods reported</p>
--	--

	<p>line 139 - the phrase "research to clinical practice pipeline" is used frequently throughout the paper. It would be useful to elaborate on figure 1 with clear definitions (i.e. is phase 2 RCTs and phase 3 pragmatic and observational studies?). Then perhaps refer to research phase.</p> <p>5. Results</p> <p>Line 222 - Reiterating previous concerns regarding including abstracts in the review based on the published study protocol. Pedantic, but 37% is not a majority. might you include a summary statement about sample sizes, such as range or mode? Presuming it is a typo but adults are defined as those over 8 years of age.</p> <p>Line 239 - Please provide a definition of scar relocation to inform interpretation of table 3.</p> <p>Line 257 - Most studies included additional objective and patient reported measures. Did any of the included manuscripts use these measures to evaluate criterion validity of US evaluation of scar thickness? If not, that would seem to be an evidence gap.</p> <p>Line 264 - From the supplementary table, the POSAS was clearly the most frequently used PROM. Perhaps give a percentage for this, i.e. of the n of studies which included a PROM, n% included the POSAS.</p> <p>Line 282 - Reliability - The data reported suggests that measuring scar thickness with US is reliable. However, it would be helpful to know more about this work to aide the interpretation of whether it is robust or there is an evidence gap. Of the 14 records that reported reliability, did the investigators report a sample size calculation to support the analysis? It would be useful to know the clinical expertise of the investigators of reliability studies. Were reliability studies conducted by radiologists, sonographers, clinicians, etc., are the results generalisable to any level of skill/training?</p> <p>Line 287 - where reporting frequencies, it would be helpful to express the value related to the n number in the total for that variable. For example, 10 studies reporting ICC is 70% of all papers reporting reliability, rather than 8% of included records.</p> <p>Line 314 - a better overview of clinical, health service, implementation and feasibility characteristics in the methods would support better interpretation of this section.</p> <p>Line 369 - It isn't clear how a 12 year longitudinal study qualifies as clinical practice</p> <p>Line 379 - Strengthens and limitations - it would be useful to provide a summary or sense of the strength of evidence here. There is only one study reporting poor correlation between US and histologically measured thickness. Is this an evidence gap, warranting further exploration?</p> <p>Line 406 - These 2 studies are in burn scar. This is an important area of work to determine minimum charge for patient reported</p>
--	---

	<p>difference in scarring of other aetiologies, so might be highlighted as an evidence gap.</p> <p>6. Discussion</p> <p>Line 426 - the authors excluded phase one studies. While there are a lack of publications regarding US use in clinical practice, this doesn't suggest it isn't happening. Perhaps conducting additional survey work would better ascertain the answer to this, and support the proposed Delphi work.</p>
--	--

VERSION 1 – AUTHOR RESPONSE

Reviewer 1 – Dr Claudia Malic:

Comment: *“The introduction is a bit long and although I appreciate for the description of USS and the problem, it should be more concised.”*

Response: We have shortened the introduction by removing a paragraph that, on review, does not contribute as strongly as other paragraphs to the overall argument of the manuscript (lines 112-132). We have also attempted to use more concise language throughout the introduction. These changes align with the comments made by Miss Kennedy (reviewer 3).

Comment: *“Another suggestion will be to use the word manuscript instead of records.”*

Response: We use the term ‘records’ throughout this manuscript as abstracts were also included in our search strategy and results. On advice from Miss Kennedy (reviewer 3), we have reported journal articles and abstracts separately throughout the manuscript where appropriate. We feel that using the word ‘manuscript’ to refer to both journal articles and abstracts may be misleading, and therefore we would prefer to continue to use the term ‘records’.

Comment: *“At the beginning of the results it was mentioned about the number of patients in all the studies but there were no further results related to this. I would prefer to see this elaborated a bit more in relations to the headings that were in the results. Although it is important how many publications addressed an item, it is also important the size of the cohorts. The search algorithm and the PRISMA should be placed as supplementary material. A table with the size of the cohorts for each subheading will be useful”.*

Response: The cohort size and sample sizes for each included record have been added to table 2 to provide context to the results. Please note that table 2 has been relocated to the supplementary files (now supplementary table 2) on request of the editorial team. On the advice of Miss Kennedy (Reviewer 3), we have included the range and mode of participants per record in the “Record Characteristics” section of the Results. The PRISMA diagram (formerly figure 2) has been moved to the supplementary files as suggested (now supplementary figure 1).

Reviewer 2 – Dr Min Yao:

Comment: *“I recommend to accept the paper for publication”.*

Response: We would like to thank Dr Yao for their confidence in the manuscript.

Reviewer 3 – Miss Donna Kennedy:

Comment: *“The search strategy, per published protocol, identified peer reviewed literature as an inclusion criteria. However, the authors have included abstracts, which account for 30% of the included references. Unless demonstrated otherwise, the assumption would be that many abstracts result from conference posters and presentations; generally not peer reviewed work, therefore should not have been included. To this point, in your abstract conclusion you note there is poor methodological reporting of included manuscripts. But this seems possibly an unfair assertion, if 30% of the body of work are abstracts with requisite abbreviated methods. This point is highlighted by the frequency of denoting ‘not reported’ in table 2. Perhaps consider if you can justify including the abstracts and if they add to or detract from the reviews take away messages. Additionally, the scope of the review is evidence across the translational pipeline, but studies in humans was in inclusion criteria. Therefore no phase one basic or pre-clinical evidence was included. Perhaps clarify the scope of the paper relative to the search strategy? Lastly, non-traumatic scarring, such as acne scarring, was a review exclusion. However, in table 1, acne scarring is used to define the “mixed scarring” category. If excluded conditions are included in the “mixed scarring” cases, does this have the potential to overinflate the total cases or patients involved in the review?”*

Response: Thank you for these insightful comments. Abstracts were included in this review to ensure that all possible ultrasound thickness measurement methods were captured and reported, and so that future studies can consider the work presented in these abstracts for completeness. We agree that

methods reported in abstracts are substantially abbreviated. We have added a statement in the discussion to reflect this (lines 978-980). While most ultrasound methods were listed as 'not reported' for abstracts, there were some abstracts that reported certain elements that we feel are important to report and add to the richness of the discussion surrounding the measurement of traumatic scars using ultrasound. We have amended our approach to using the data from abstracts as suggested, to substantiate the main findings (i.e., whether they add to or detract from the reviews take-away messages). Stratification of peer-reviewed journal articles and abstracts in the results are intended to make it easy for the reader to follow this approach. The reasoning behind including abstracts in the results is now signposted in the "record selection" section of the methods (lines 196-202).

The inclusion criteria used in this review is indeed evidence across the translational pipeline, however the inclusion criteria are limited to studies in humans. This would preclude phase 1 or preclinical trials from inclusion. A statement has been added to the manuscript to reflect that phase 1 studies are not included in the review (lines 224-228).

Acne scarring was only included under 'mixed scarring' in this review where traumatic scars were also measured, not when acne scars were the focus. Where traumatic scars are being measured alongside other scarring types (e.g., acne scarring), and no further detail is provided, it is assumed that the same methods are used, which can therefore be included in the review.

Comment: *"The data extraction fields reported in supplementary table 1 are comprehensive, and it would be helpful to have more of this data included in the manuscript to improve interpretability. For example, in table 2, perhaps list included studies alphabetically but by study phase. Additionally, include the country where the work was completed, funding sources (i.e., commercial or non-commercial) and where scar aetiology is surgical, reporting the surgical cohort or condition. This will enable the reader to better appreciate gaps and challenges to progressing US to fruition as a clinical measurement tool and may promote collaboration and cross-fertilization, as appears to be the overarching remit of the review."*

Response: Two columns have been added to supplementary table 2 (formerly table 2) to include the countries where the included works were conducted and the funding source for each record. Additional information relating to the funding source of each publication was extracted. This has been updated in supplementary table 1. The type of surgical intervention has also been specified where the

reported the scar aetiology was “surgical” (supplementary table 2). We agree that this information will help promote collaboration and cross-fertilization. Supplementary table 2 has been amended following suggestions to list records alphabetically by author within each phase of the translational pipeline. A summary table has been created and replaced the original table 2 in the manuscript. This reports numbers of records per pipeline phase.

Comment: *“Abstract, line 33. Methods are reported in the literature. The abstract results and conclusion appear to omit important findings, for example reliability data. The conclusion centres on the authors’ recommendations around standardising measurement, whereas my take on this is that the authors have extracted information to support standardisation and curated supplementary table 4.”*

Response: Thank you for your feedback. We have re-worked portions of the abstract to improve the abstract in line with the comments. We have included a brief statement on reliability (lines 51-52) and have rewritten the end of the conclusion to provide a more accurate overall picture of the research conducted. Given the 300-word limit of the abstract it was difficult to reflect all of the results but we hope we have now presented a more complete picture of the findings.

Comment: *“Introduction”*

“Line 78 – how is a “major” skin injury defined?”

Response: This phrase has been removed based on review comments provided by Reviewer 1 (Dr. Claudia Malic) to make the introduction more succinct (lines 85-87).

Comment: *“Line 98 – While this is interesting, the relevance to of the skills required to perform and interpret ultrasound as related to the review aims is not clear.”*

Response: As with line 78 above, this section has been removed on the advice provided by Reviewer 1. We agree that this section detracts from the overall narrative of the review.

Comment: *“Line 124 – This reads as though US has successfully been implemented to support personalised medicine. Perhaps elaborate on what treatment protocols you are referring to? Does this apply to all scars and all interventions?”*

Response: We agree that the terminology used in this section is confusing. To that end, we have changed the wording of this section to reflect that we are outlining the potential for ultrasound in personalised medicine (lines 139). To our knowledge, this potential is yet to be realised.

Comment: *“Line 132 – Methods reported”*

Response: On the advice of Reviewer 3, this sentence has been reworded to improve the readability of the manuscript (line 152).

Comment: *“Line 139 – The phrase “research to clinical practice pipeline” is used frequently throughout the paper. It would be useful to elaborate on figure 1 with clear definitions (i.e., is phase 2 RCTs and phase 3 pragmatic and observational studies?). Then perhaps refer to research phase.”*

Response: We thank you for your comment and agree that further detail should be provided. As such, pipeline phases are now defined with greater detail in the methods section (lines 224-228).

Comment: *“5. Results*

Line 222 – Reiterating previous concerns regarding including abstracts in the review based on published study protocol. Pedantic, but 37% is not a majority. Might you include a summary statement about sample sizes, such as range or mode? Presuming it is a typo but adults are defined as those over 8 years of age.”

Response: As mentioned above, we believe that the inclusion of abstracts in this review is warranted, and adds to the richness of the data and potential for future reviews to build on these findings. We feel that it is important to include these as, prior to the review being conducted, we were unsure whether any additional information may be presented in these abstracts that were not included in full-text articles. We have now presented results from peer-reviewed journal articles and abstracts separately throughout the manuscript to better explain the data. We have changed the wording of this section to reflect that 37% is not a majority, and have fixed the typo to define adults as aged older than 18 (lines 264-266). We have included additional descriptive statistics relating to the sample sizes to better explain the data. This includes both range and mode (lines 251-252).

Comment: *“Line 239 – Please provide a definition of scar relocation to inform interpretation of table 3.”*

Response: A definition of scar relocation is provided in line 588. We appreciate that this may be confusing, and have added a definition to the footnotes of table 3. Please note that table 3 has now been moved to the supplementary files (supplementary table 3) following editorial requests.

Comment: *“Line 257 – Most studies included additional objective and patient-reported measures. Did any of the included manuscripts use these measures to evaluate criterion validity of US evaluation of scar thickness? If not, that would seem to be an evidence gap.”*

Response: Thank you for these insightful comments. We have included a statement in the results (lines 391-406) and discussion (lines 576-587 and lines 906-910) addressing this issue. While this is not within the purview of this investigation, it is an important factor that we have now highlighted.

Comment: *“Line 264 – From the supplementary table, the POSAS was clearly the most frequently used PROM. Perhaps give a percentage for this, i.e. of the n studies which included a PROM, n% included the POSAS.”*

Response: A statement has been added which presents the number of records reporting the use of the POSAS (lines 567-576).

Comment: *“Line 282 – Reliability – The data reported suggests that measuring scar thickness with US is reliable. However, it would be helpful to know more about this work to aide the interpretation of whether it is robust or there is an evidence gap. Of the 14 records that reported reliability, did the investigators report a sample size calculation to support the analysis? It would be useful to know the clinical expertise of investigators of reliability studies. Were reliability studies conducted by radiologists, sonographers, clinicians etc., are the results generalisable to any level of skill/training?”*

Response: Regarding sample sizes for reliability estimates, in most cases, reliability was calculated as secondary outcome measures. Sample size calculations for these studies were therefore conducted for primary outcome measures, with reliability information provided for context. The clinical expertise of the investigators where reliability was calculated have been added to the manuscript (lines 676-681), with an additional statement made in the discussion relating to the generalisability of the results of these studies (lines 966-971).

Comment: *“Line 287 – where reporting frequencies, it would be helpful to express the value related to the n number in the total for that variable. For example, 10 studies reporting ICC is 70% of all papers reporting reliability, rather than 8% of included records”.*

Response: Reported percentages throughout the manuscript have been amended to reflect journal articles and abstracts. Where appropriate, percentages have been recalculated to reflect the number reported for each variable, rather than all included records.

Comment: *“Line 314 – a better overview of clinical, health service, implementation and feasibility characteristics in the methods would support better interpretation of this section”.*

Response: An overview of the clinical, health service, implementation and feasibility characteristics are presented in the methods section (lines 230-243), and outlined in supplementary table 1. Some definitions of the most commonly reported implementation outcomes in this review have now been defined in the methods section (lines 234-237).

Comment: *“Line 369 – It isn’t clear how a 12 year longitudinal study qualifies as clinical practice”*

Response: This study was determined to be a phase 4 study investigating clinical practice as pressure garments were stated to be used in regular clinical practice at the study location, thus was deemed to relate to clinical practice over that long period.

Comment: *“Line 379 – Strengthens and limitations – it would be useful to provide a summary or sense of the strength of evidence here. There is only one study reporting poor correlation between US and histologically measured thickness. Is this an evidence gap, warranting further exploration?”*

Response: The overall aim of the scoping review was to summarise the methods used to measure traumatic scar thickness using ultrasound. We did not specifically address the strength or quality of the evidence as is common in scoping reviews. In line with our aim, this section of the review summarised the need for holistic measurement. As suggested, we have added a statement about the poor correlation between US and histologically measured thickness being worthy of further exploration.

Comment: *“Line 406 – These 2 studies are in burn scar. This is an important area of work to determine minimum change for patient reported difference in scarring of other aetiologies, so might be highlighted as an evidence gap.”*

Response: We agree that the studies mentioned above are both in burn scar, and further work is required to determine whether these results can be generalised to other scar aetiologies. We have included a small statement (lines 869-876) to reflect this. To our knowledge, no further investigations have been conducted to identify the correlation between ultrasound measurements and patient-reported differences in scar thickness.

Comment: *“6. Discussion*

Line 426 – the authors excluded phase one studies. While there are a lack of publications regarding US use in clinical practice, this doesn't suggest it isn't happening. Perhaps conducting additional survey work would better ascertain the answer to this, and support the proposed Delphi work.”

Response: We agree and have expanded the statement in the discussion relating to the lack of phase 4 studies (lines 894-910). Further investigation, possibly through surveys of team practices, is likely required to further understand the ultrasound methods used in routine clinical practice that are not reported in research articles.

Once again, we would like to thank all three reviewers for their constructive comments in improving the quality of our manuscript.

Kind regards,

Brandon Meikle (on behalf of the authorship team)

PhD Candidate

Centre for Children's Burns and Trauma Research (CCBTR)

The University of Queensland Child Health Research Centre

Centre for Children's Health Research

Level 7, 62 Graham Street, South Brisbane, QLD, Australia, 4101

Faculty of Medicine, The University of Queensland, Australia

E: brandon.meikle@uq.net.au

VERSION 2 – REVIEW

REVIEWER	Kennedy, Donna Imperial College Healthcare NHS Trust, Therapy Department
REVIEW RETURNED	16-Feb-2024

GENERAL COMMENTS	<p>Thank you for the opportunity to review the revision of this manuscript. The authors are to be commended for producing a review that will inform research agendas and practice developments. The synthesis of data reported in tables 2 and 3 provides an excellent, relevant and accessible summary of the included literature.</p> <p>Minor typographical errors for editing: Line 207; Table 1. Please correct phrasing: Measurement of non-traumatic scars (e.g., acne scars). Where non-traumatic scars were measured along with burn scars, these were included.</p> <p>Line 234 Acceptability is defined as is defined as the level to which ultrasound.... "Is defined as" is repeated.</p> <p>Line 264 Adults aged 18 years and older were most commonly targeted in articles was adults aged 18 years and older. Rephrase please.</p> <p>Line 410 B-mode, including high-frequency (i.e., ≥ 20 MHz) B-mode ultrasound was the most 410 commonly reported ultrasound type used in both articles I believe both should be most?</p>
-------------------------	--

VERSION 2 – AUTHOR RESPONSE

Reviewer 3 – Dr Donna Kennedy:

Comment: *"Line 207; Table 1. Please correct phrasing: Measurement of non-traumatic scars (e.g., acne scars). Where non-traumatic scars were measured along with burn scars, these were included."*

Response: The phrasing in this table has been corrected (Table 1, line 151) as per Dr Kennedy's suggestion above.

Comment: *"Line 234: Acceptability is defined as is defined as the level to which ultrasound... - "Is defined as" is repeated."*

Response: This was an error which has been corrected to now read reads “acceptability is defined as the level...” (line 176).

Comment: “Line 264: Adults aged 18 years and older were most commonly targeted in articles with adults aged 18 years and older – rephrase please.”

Response: Again, this was an error submitted in the marked copy of the manuscript which has been corrected in lines 204-205 to now read “Adults aged 18 years and older were the most highly represented age group reported in articles (n = 43 articles; 52% of articles)...”

Comment: “Line 410: B-mode, including high-frequency (i.e., ≥ 20 MHz) B-mode ultrasound was the most commonly reported ultrasound type used in both articles – I believe both should be most?”

Response: The word “both” has now been removed from this sentence to improve the clarity. This sentence now reads “B-mode, including high frequency B-mode ultrasound (i.e., ≥ 20 MHz) was the most commonly reported ultrasound type in the included articles...” (Line 227).

Kind regards,

Brandon Meikle (on behalf of the authorship team)

PhD Candidate

Centre for Children’s Burns and Trauma Research (CCBTR)

The University of Queensland Child Health Research Centre

Centre for Children’s Health Research

Level 7, 62 Graham Street, South Brisbane, QLD, Australia, 4101

Faculty of Medicine, The University of Queensland, Australia

E: brandon.meikle@uq.net.au

P: (07) 3039 7393