Author's response to reviews

Title: Level of agreement between patient-reported EQ-5D responses and EQ-5D responses mapped from the SF-12 in an injury population

Authors:

Emma McDermott (emma.mcdermott@monash.edu)
Pamela M Simpson (pamela.simpson@monash.edu)
Sarah Derrett (sarah.derrett@otago.ac.nz)
Shanthi Ameratunga (s.ameratunga@auckland.ac.nz)
Suzanne Polinder (s.polinder@erasmusmc.nl)
Ronan A Lyons (r.a.lyons@swansea.ac.uk)
Frederick P Rivara (fpr@u.washington.edu)
James E Harrison (james.harrison@flinders.edu.au)
Belinda J Gabbe (belinda.gabbe@monash.edu)

Version: 3
Date: 8 January 2015

Author's response to reviews: see over
7th January 2015

Kate Muller
Editorial Team
Population Health Metrics

Dear Ms Muller,

Re: "Level of agreement between patient-reported EQ-5D responses and EQ-5D responses mapped from the SF-12 in an injury population?"

On behalf of my co-authors, I am pleased to submit this revised version of our manuscript for consideration by Population Health Metrics. Each of the reviewers’ comments has been addressed as described in the following pages, and the changes to the manuscript are shown as track changes.

We look forward to your decision on this paper. Should you have any queries regarding this manuscript, please do not hesitate to contact me on +61 3 9903-0951 or by email (belinda.gabbe@monash.edu).

Yours sincerely

Prof Belinda Gabbe
Head, Prehospital, Emergency and Trauma Research
NHMRC Career Development Fellow
Response to Reviewers’ comments

Reviewer 1

Thank you for inviting me to review this interesting study. The manuscript is concise and well written. The subject of the paper is well defined and new in the area of injury. Strengths of this paper are empirical data from a large sample size with expected high prevalence of disability (patients with severe injury (ISS>15) and elderly admitted to hospital because of orthopaedic trauma). I do have a few questions and suggestions for improving the paper.

Major revisions

1) The supplementary material that shows the results of the analysis for VOTOR and VSTR separately are interesting. However, I wonder if the effect of more severe injury (VSTR) and the effect of age (VOTOR) on HRQL measured by EQ-5D and SF-12 may partly cancel each other out. Do you think this may have had an effect?

Author response: The supplementary table show no difference in the level of agreement for each item of the EQ-5D by the source of data. The VSTR and VOTOR studies represent different injury patient groups. VSTR patients are more severely injured than in VOTOR while those in VOTOR are older. Both age and severity would make the HRQL worse at follow-up. So while the results of the two databases are similar, the reasons for this are different. The fact that the level of agreement was similar despite the differences in the patient populations studied adds further strength to the findings. Further explanation has been added to the Discussion of the revised manuscript to highlight this point.

2) I suggest to include an additional analysis to investigate if there are differences in agreement by ISS category (VSTR) e.g. ISS 16-25 and >25? Or maybe it is better to use the VOTOR if the registry includes a variable that registers severity of the injury?

Author response: While VOTOR does not collect ISS scores specifically, VOTOR excludes any case meeting VSTR criteria and therefore represents a less severely injured patient group (i.e. ISS<15). The supplementary table shows that the level of agreement was not different between the studies. The fact that the Injury-VIBES VOTOR dataset excludes any orthopaedic trauma cases meeting VSTR criteria has been added to the Methods section of the revised manuscript. Given the findings were similar for VOTOR and the VSTR, and the results are shown together in the manuscript and separately in the supplementary material, excluding the VSTR would be unnecessary. The study is focused on disability while the ISS is a threat to life measure. The cut-offs suggested relate to threat to life rather than threat to disability and are not relevant to the current study. There is no ISS equivalent for disability.

Minor revisions

Introduction

3) The authors state that both the EQ-5D and SF-12 are commonly used in injury populations. I wonder if this is true. From Figure 1 in the Polinder review that is referred to by the authors and that included 41 articles there appear to be only three studies that used SF-12 and 5 studies that used EQ-5D to assess HRQL.

Author response: The reviewer’s comment is acknowledged. The Polinder review did not include all injury studies which used the EQ-5D or SF-12 instruments as its focus was general injury studies rather than specific types of injury (e.g. traumatic brain injury, burns, etc). Overall, according to Ovid Medline, the SF-36 has been used in 806 injury or trauma studies, the EQ-5D has been used in 199, the SF-12 in 129, the HUI in 27, the QWB in 18 and the WHODAS in 12. These search results support the premise that the EQ-5D and SF-12 instruments are used commonly in injury studies. Not all of these references could be added but additional references have been added to the Introduction to support the statement. Of note, the SF-36 is used more commonly than the SF-12 and EQ-5D and this has been acknowledged in the revised manuscript.

Methods

4) This section is clear and explains the study design and measures very well. The EQ-5D and SF-12 data were collected at 6 and 12 months post-injury. Were the EQ-5D and SF-12 presented in the same order?
Author response: The EQ-5D and SF-12 were presented in the same order at both time points. This has been added to the Methods section of the revised manuscript.

5) In the VSTR there may be patients with severe cognitive problems that may have caused inconsistencies in the answers to the EQ-5D and SF-12 questions. How did the authors take this into account?

Author response: The issue of cognitive issues was not confined to the VSTR. The higher age of VOTOR patients was associated with a higher prevalence of cognitive issues such as dementia. Unlike the EQ-5D, there is no proxy version of the SF-12, and therefore the SF-12 is not administered where the interview cannot be conducted directly with the patient. As this study required patients to have both SF-12 and EQ-5D responses, these patients were excluded from the study. This has been clarified in the Methods section of the revised manuscript and also noted as a limitation of the study (and the mapping process overall) in the revised Discussion section.

Results

6) Number of patients that were included in the analysis: I don’t understand the explanation of the total number of patients at each of the follow up points. “….7504 patients had data available at 6-months only, 8722 patients had 12-month data only, and 6060 patients had data at both time points”. 7504 + 8277 + 6060 = 21841 patients, but the authors report 10166 patients were included in the study. Could you please explain this?

Author response: We apologise for the confusion. There were 10166 patients overall; 6060 had data at both time points, 2662 had data at 12 months but not 6 months, and 1444 had data at 6 months but not 12 months. This has been clarified in the Results section of the revised manuscript.

7) How many patients were included from the VOTOR and how many from the VSTR?

Author response: These figures are included in the Supplementary Table.

8) Table 1: cause of injury. The causes of injury seem to be a mixture of mechanism of injury (fall) and type of road user (e.g. motor vehicle). I suggest changing the road users to “road user accident” (e.g. motor vehicle accident).

Author response: Road user is an overarching term that includes multiple causes of injury. We have amended the table to address the reviewer’s comment but have not used “accident” but crash or incident.

9) Table 2: light shading is missing in the cell 12 months, mobility, some problems (actual)/severe problems (estimated).

Author response: Amended as suggested.

Discussion and conclusions

The discussion and conclusions well balanced and adequately supported by the data.

Author response: No changes required.

10) Do the authors have any recommendations for future research in this area?

Author response: Recommendations for further research are included in the conclusions section of the manuscript.
Reviewer 2

Well written manuscript that tries to answer the question whether or not EQ5D (utility) scores can be estimated from SF-12 scores, as a validation of mapping algorithms. Main conclusion is that the efforts yield moderate to substantial agreement, and that disability was consistently underestimated.

1. Is the question posed by the authors new and well defined?
Yes, the question is original and very relevant in burden of disease research, for example.

Author response: No changes requested.

2. Are the methods appropriate and well described, and are sufficient details provided to replicate the work?
Although the methods are described appropriately, the actual algorithms to convert SF-12 scores into EQ5D scores have not been described. Perhaps these could be added to the manuscript as supplementary material.

Author response: The reference Gray et al (2006) clearly describes the algorithm for converting the SF-12 scores into EQ-5D scores. We do not believe that reproducing their algorithm is needed and the basis of the algorithm is already described in the Methods section.

Furthermore, the authors describe the Injury VIBES patient sample as hospitalized major trauma patients who met any of three criteria: 1) ISS > 15, 2) ICU admission >24 hrs, and 3) requiring urgent surgery between 2007 and 2013. Criteria one and three are not very well specified.

Author response: The ISS is a standardised injury severity measure that is used globally. An ISS>15 is a routinely used cut-off to define major trauma and we have added an appropriate reference for the ISS. We acknowledge that the urgent surgery criterion is not well defined and this has been addressed in the revised manuscript. The VSTR defines urgent surgery as within 24 hours of injury that involve intra-cranial, intra-thoracic or intra-abdominal operations, or fixation of pelvic or spinal fractures. This definition, and the appropriate reference, has been added to the Methods section of the revised manuscript.

Initially, it was not clear to me that both instruments were administered at both moments in time (6 and 12 mths).

Author response: It is a strength of our study that both the EQ-5D and SF-12 were administered at the same time. We have amended the last sentence of the Datasets and Participants paragraph in the Methods to further clarify this.

3. Are the data sound and well controlled?
The authors give a certain value judgment over the range of Kappa scores (poor agreement, slight, fair, moderate, substantial and almost perfect agreement) based on Landis and Koch (1977), but more recent literature (Gwet, 2010; Handbook of Inter-Rater Reliability, Second Edition) state that this set of guidelines is not universally accepted; they supplied no evidence to support it but personal opinion.

Reviewer response: There is no universally accepted guideline for interpreting Kappa. Gwet has developed alternative models for inference but these have not been adopted and the Landis and Koch guidelines remain the most commonly used which is why these were included, and referred to. Gwet supports the use of the Landis and Koch guidelines over other commonly used methods such as the Fleiss guidelines. The rationale for using the Koch and Landis guidelines has been added to the Methods and additional references added. It should also be noted that we have not only presented the Kappa coefficients, but have also used prevalence and bias adjusted Kappa coefficients as well as the percentage agreement to ensure a more comprehensive summary of agreement.

Furthermore, regarding table 4, what might be more informative than the mean difference between actual disability weight and estimated disability weight (with positive and negative differences balancing out) , is to have insight into the means of the absolute differences for each of the EQ5D dimensions.
Author response: The absolute differences (and corresponding 95% CI) between the EQ-5D preference weights have been added to Table 4 as suggested.

Perhaps an expert statistician could share his/her views on these issues.

Author response: We would be happy for the manuscript to undergo a statistical review. The author list includes multiple biostatisticians and we are confident that the statistical analysis is appropriate and comprehensive.

4. Does the manuscript adhere to the relevant standards for reporting and data deposition?

I do not have enough information to make a sound judgement on that. One of the things that did strike me is the relatively old literature. Only 3 of the 29 references used is from three years ago or more recent.

Author response: The reference list has been further updated with the revised manuscript.

As the data are derived from the VOTOR and VSTR databases, what involvement does Injury-VIBES have in this manuscript?

Author response: The study was conceived by the Injury-VIBES investigators and forms a component of the original Injury-VIBES protocol. The Injury-VIBES investigators developed the study, the analysis plan and the interpretation of the findings.

5. Are the discussion and conclusions well balanced and adequately supported by the data?

The authors do not give a clear cut answer to the research question, they leave room for interpretation, while, to my opinion, the data would lead to the conclusion that the mapping from SF-12 data to EQ5D estimations would not be robust enough to use in research (as far as I have an understanding of the kappa statistic).

Author response: The conclusion in both the Abstract and main manuscript clearly state that caution should be taken when mapping from the SF-12 to the EQ-5D in injury studies, and the reason for the need for caution is stated.

6. Do the title and abstract accurately convey what has been found?

Actually, the conclusion in the abstract leaves less room for interpretation than the conclusion section in the manuscript (as described above): caution should be taken! The title might be amended to reflect (part of) the conclusion.

Author response: See previous response. The title states what the study was about rather than the findings. “Level of...” has been added to the beginning of the title to ensure that there is no indication that agreement exists, only that the study is assessing the level of agreement.

7. Is the writing acceptable?

Apart from a few very long sentences (e.g., introduction: last sentence of first paragraph; penultimate sentence of third paragraph), yes.

Author response: The manuscript has been further edited to avoid the long sentences noted by the reviewer.