PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Comparison of the EQ-5D-3L and the SF-6D (SF-36) contemporaneous utility scores in patients with Chronic Kidney Disease in Sri Lanka; a cross sectional survey
AUTHORS	Kularatna, Sanjeewa; Senanayake, Sameera; Gunawardena, Nalika; Graves, Nicholas

VERSION 1 – REVIEW

REVIEWER	Shunping Li School of Health Care Management, Shandong University, China
REVIEW RETURNED	14-Jul-2018

GENERAL COMMENTS	Generally, this manuscript is well-structured and this topic is of importance for health utility measurement. However, some concerns should be addressed before it is considered for publication. Introduction: 1. In introduction section, the purpose of this study is to compare contemporaneous EQ-5D-3L and SF-6D utility scores in patients with CKD, I think it should be concluded that which tool is recommended, or which one is preferred for CKD patients based on the evidences in this paper. 2. Since there is a tariff in Sri Lanka about EQ-5D, it should be used. The purpose is to compare the two measures, and it would be used in Sri Lanka in the future, so it is appropriate to use the country-specific tariff here. Methods: 3. The date of data collection should be added in this section. 4. The normality should be tested before t-test. It should not be illustrated only by graphs. 5. It should be described that this is a face-to-face survey, not just mentioned in the final limitations, and also give readers more details about the survey. 6. For SF-36, the response levels are from two to seven, I think it should be 2-6. Could you explain it or make sure of this? 7. Since the purpose of this study is to compare, why don't the authors use a Bland-Altman plot to assist it. 8. The calculation formula of “effect size” should be explained clearly, or at least give one or more references. 9. Why do the authors only use Spearman correlation coefficient in agreement analysis? ICC should be used here. Results and discussions: 10. Table 1, I did not find the meaning of subgroup to compare the utility means of EQ-5D and SF-6D. I think it will be useful to compare each tool within subgroups. 11. The results of test-retest reliability is too simple. The two specific coefficient should be indicated and not just show that it is more than 0.9 directly. Actually, there should be two coefficients for the two

	tools. 12. The ceiling effect is a little lower than most other studies and this should be discussed in the discussion section. 13. In conclusion, there should be which one is better and the two tools can be interchangable. 14. The F value of variance analysis be not displayed in the table 5. Format: 15. In the part of discussion, the fourth paragraph, “Though overall ceiling and floor effects of both instruments were small, significant ceiling effect was evident in the EQ-5D-3L. This was consistent with several other studies conducted elsewhere” The ceiling effect of EQ-5D-3L is 11.8% (<15%). This sentence is ambiguous, could the authors make it more clear? 16. “At present, there is no consensus on the methodology to compare the utility scores of different MAUIs” ---please add citations. 17. In the table 1, the format of the number “07’ is not standardized. And it would be better to mark the p values if the p<0.05 in table1. 18. Glomerular Filtration Rate (GFR) was less than 60 ml/min per 1.73m2 of body surface area…. The number “2” should be superscripted. 19. The middle spacing of the article should be uniform.

REVIEWER	Susan Stewart National Bureau of Economic Research, United States
REVIEW RETURNED	03-Aug-2018

GENERAL COMMENTS	It is interesting to compare these two instruments in a specific disease population. However the implications of the results are not not clear. How would the choice of instrument affect results of cost-effectiveness analysis or patient care? The authors mention in the objectives in the abstract and on page 5 in the introduction that the choice of instrument "could adversely influence the results of CUA and thereby the decision-making process". However, they do not discuss the results of their study in that context. The authors conclude that the two tools cover different 'spaces in health'. While this is not quite grammatically correct, their key finding is that the instruments cover different aspects of health. Summarizing these differences and discussing their implications is important. Many statistical results are presented, and if is difficult for the reader to determine overall how these results should guide their choice of instrument or their interpretation of studies using the different instruments. The authors state that because the utility assessment was done only once, the 'responsiveness' of the instruments was not assessed (second line on page 11). I think what is meant is reliability? Some other smaller grammatical changes are needed. For example on page 5 line 17, "differences of different" should be differences between". page 11, line 19, the statement "but different to several other studies" should be "different from other studies". There are also some places in the article where repetition could be eliminated (such as the first paragraph under 'agreement and differences' on page 7). In the abstract, the second-last sentence of the results section should ideally be re-worded to avoid any accidental implication that this is a longitudinal study. (Instead of saying that the scores

	decreased with advancing CKD stage, could say that they were significantly lower for those in more advanced stages.) When describing instruments' scoring, 1.00 on the EQ-5D is said to indicate 'good health status' (page 7, line 7). More accurately, it is meant to represent 'best imaginable health'. (The following sentence, stating that values close to zero indicate worse conditions while 1 represents perfect health, is repetitive and could be omitted.) To be consistent when describing the SF-6D, I suggest stating that its scores range from 0.296 to 1.0 (giving lowest value first). Under 'Discrimination' on page 8, the first sentence (line 11) should state that MAUIs are meant to measure change in QOL (rather than just improvement), reflecting change in health among those with the condition of interest. It is good that at the end of page 11 the authors acknowledge that the different disease stages may not actually always represent different states of health-related quality of life. That is an important reason for measuring HRQOL, and could be mentioned in the other areas of the paper where discrimination is discussed. Our 2014 study seems relevant to cite: Stewart, S. T., Cutler, D.M. & Rosen, A.B. (2014). Comparison of Trends in U.S. Health-Related Quality of Life over the 2000's Using the SF-6D, HALex, EQ-5D, and EQ-5D Visual Analog Scale versus a Broader Set of Symptoms and Impairments. Medical Care, 52(12), 1010-1016.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1	
Introduction:	
1. In introduction section, the purpose of this study is to compare contemporaneous EQ-5D-3L and SF-6D utility scores in patients with CKD, I think it should be concluded that which tool is recommended, or which one is preferred for CKD patients based on the evidences in this paper.	Added the following sentence to the conclusion. Line 417-420 Finally, SF 6D had the lowest floor and ceiling effect, and was better at detecting different stages of the disease. Thus based on the evidence presented in this study, SF 6D appears to be more appropriate to be used among CKD patients
2. Since there is a tariff in Sri Lanka about EQ-5D, it should be used. The purpose is to compare the two measures, and it would be used in Sri Lanka in the future, so it is appropriate to use the country-specific tariff here.	Kindly note that this is the first known study which compared the utility scores of EQ-5D-3L and SF-6D among CKD patients. Thus the main aim of this study was to assess how these scores differ among CKD patients, which will guide the researchers the appropriateness of using the 2 tools during economic evaluations among CKD patients, which may require prospective collection of utility data.

	If we had used EQ-5D-3L Sri Lankan utility scores, we would be comparing Sri Lankan EQ-5D-3L utility scores with UK SF 6D utility scores. It is found that the utility scores of the same instrument across different countries vary considerably [(1)]. Kularatna et al. indicate that the Sri Lankan EQ-5D-3L utility values show marked differences to the other countries, especially for the more severe health states [(2)]. Thus if we use utility scores of two different instruments of two different countries, we believe the comparison will be useless. Thus we believe the use of UK based utility scores of both the tools is appropriate in this study. *[1] Knies S, Evers SM, Candel MJ, Severens JL, Ament AJ. Utilities of the EQ-5D. Pharmacoeconomics. 2009 Sep 1;27(9):767-79.* *[2] Kularatna S, Whitty JA, Johnson NW, Jayasinghe R, Scuffham PA. Valuing EQ-5D health states for Sri Lanka. Quality of Life Research. 2015 Jul 1;24(7):1785-93.*
Methods:	
3. The date of data collection should be added in this section.	Following sentence was change in the methods section. Line 122 "A population-based descriptive cross-sectional study was conducted in the district of Anuradhapura in the North Central Province (NCP) of Sri Lanka between Septembers to December 2015".
4. The normality should be tested before t-test. It should not be illustrated only by graphs.	Thank you for pointing this. Results indicated that the utility scores were non-normally distributed. Thus, non-parametric tests were used to find the significance of the difference. Changes were done appropriately to the text and to the tables. The following section was added to the methods (line 194-196)

	"Normality of the two distributions were assessed using Kolmogorov-Smirnov and Shapiro-Wilk tests. Wilcoxon signed-rank test was used to assess the difference between the two instruments in each socio-demographic class." Further changes were done to the results section and tables depending on the test outcome. Line 291-299.
5. It should be described that this is a face-to-face survey, not just mentioned in the final limitations, and also give readers more details about the survey.	The following section was added to the methods; (Line 135 – 143) "The study instrument was an interviewer-administered questionnaire to gather information on the socio-demographic information, CKD related information, EQ-5D-3L and SF 36. Five Public Health Inspectors working in the CKD unit in the North Central Province were used for the data collection and all have been working in the unit for more than 5 years and they had experience in functioning as data collectors for many local and international studies done among CKD patients in the NCP. Data collection was mostly done on weekdays considering the fact that most of the study units were expected to be at home, since most are employed in the informal sector. "
6. For SF-36, the response levels are from two to seven, I think it should be 2-6. Could you explain it or make sure of this?	Thank you for pointing the mistake. It is 2-6. Corrected appropriately. Line 174
7. Since the purpose of this study is to compare, why don't the authors use a Bland-Altman plot to assist it.	Thank you for the suggestion. Bland-Altman graph added.
8. The calculation formula of "effect size" should be explained clearly, or at least give one or more references.	Another reference added

9. Why do the authors only use Spearman correlation coefficient in agreement analysis? ICC should be used here.	Thank you for the suggestion. Change made. Line 223, 258 Further changes were done to the results section depending on the test outcome. Line 311-324.
Results and discussions:	
10. Table 1, I did not find the meaning of subgroup to compare the utility means of EQ-5D and SF-6D. I think it will be useful to compare each tool within subgroups.	Thank you for the suggestion. But kindly note that the main aim of this study was to see how the utility scores of the two instruments differ in different situations. Eg: In both males and females the utility scores of the 2 instruments differ significantly Thus we believe the best method to demonstrate how the utility scores of the two instruments differ, is by the way the table 1 is presented.
11. The results of test-retest reliability is too simple. The two specific coefficient should be indicated and not just show that it is more than 0.9 directly. Actually, there should be two coefficients for the two tools.	Correction made. Line 338
12. The ceiling effect is a little lower than most other studies and this should be discussed in the discussion section.	The following section was added to the discussion (385-392) "Further, the current study reported relatively lower ceiling effect, for the EQ 5D, compared to results obtained among Parkinson's disease (13.5%) and stable angina (15.5%) patients. However, our result was higher compared to the ceiling effect observed among patients with systemic sclerosis (7.0%). Among many other factors that could contribute to these differences, the level of morbidity of a disease is said to be one of the factors which could influence the ceiling effect observed in EQ 5D (50). Thus the diseases with lower morbidity are expected to have higher ceiling effects."

13. The F value of variance analysis be not displayed in the table 5.	Correction made to the table 5
14. In the conclusion there should be which one is better and the two tools can be interchangeable.	Changes made to the conclusion (line 417-420)
15. In the part of discussion, the fourth paragraph, “Though overall ceiling and floor effects of both instruments were small, significant ceiling effect was evident in the EQ-5D-3L. This was consistent with several other studies conducted elsewhere” The ceiling effect of EQ-5D-3L is 11.8% (<15%). This sentence is ambiguous, could the authors make it more clear?	Changed the sentence. Line 379-382 “Though overall ceiling and floor effects of both instruments were small, relatively higher ceiling effect was evident in the EQ-5D-3L. This was consistent with several other studies conducted elsewhere, where EQ 5D 3L reported a relatively higher ceiling effect compared to SF 6D”.
16. “At present, there is no consensus on the methodology to compare the utility scores of different MAUIs” ---please add citations.	Reference added
17. In the table 1, the format of the number “07’ is not standardized. And it would be better to mark the p values if the p<0.05 in table1.	Correction made to the table 1
18. Glomerular Filtration Rate (GFR) was less than 60 ml/min per 1.73m2 of body surface area…. The number “2” should be superscripted.	Corrected. Line 126
19. The middle spacing of the article should be uniform.	Corrected
Reviewer 2	
1. It is interesting to compare these two instruments in a specific disease population. However the implications of the results are not not clear. How would the choice of instrument affect results of cost-effectiveness analysis or patient care? The authors mention in the objectives in the abstract and on page 5 in the introduction that the choice of instrument "could adversely influence the results of CUA and thereby the decision-making process". However,	Thank you for pointing this out. The following section was added to the discussion (Line 349-354) “Evidence indicate that the choice of MAUI (Eg; EQ 5d or SF6D) has an impact on the results of the cost-utility analysis (45, 46). Sack et al. (2009) compared the results of cost-utility estimates using both EQ 5D and SF 6D. Results

they do not discuss the results of their study in that context.	indicated contrasting results for the two instruments and authors concluded that the choice of the instrument does matter in cost-utility analysis (45). Thus, from an economic perspective it is important to know the most suitable MAUI to be used among CKD patients." Line 417-420 "Finally, SF 6D had the lowest floor and ceiling effect, and was better at detecting different stages of the disease. Thus based on the evidence presented in this study, SF 6D appears to be more appropriate to be used among CKD patients."
2. The authors conclude that the two tools cover different 'spaces in health'. While this is not quite grammatically correct, their key finding is that the instruments cover different aspects of health. Summarizing these differences and discussing their implications is important. Many statistical results are presented, and if is difficult for the reader to determine overall how these results should guide their choice of instrument or their interpretation of studies using the different instruments.	Wording changed in the 'spaces in health'. Line 415 The conclusion was changed to summarize the findings and guide the researches to choose the best tool to be use among CKD patients.
3. The authors state that because the utility assessment was done only once, the 'responsiveness' of the instruments was not assessed (second line on page 11). I think what is meant is reliability?	Responsiveness of a MAUI is determined by its ability to capture health-related quality of life changes over time (1). Measuring the kidney function (eGFR or CKD stage) over a time period (eg: one year apart) and assessing the response of the tool to the change in the kidney function over time would have been the ideal way of measuring the responsiveness in the current study. But unfortunately only one time assessment was done among the study population. Whereas, reliability is the overall consistency of a measure over time. Thus the two measures measure different aspects of the

	instruments. (1) *Kularatna, S., Byrnes, J., Chan, Y.K., Ski, C.F., Carrington, M., Thompson, D., Stewart, S. and Scuffham, P.A., 2017. Comparison of the EQ-5D-3L and the SF-6D (SF-12) contemporaneous utility scores in patients with cardiovascular disease. Quality of Life Research, 26(12), pp.3399-3408.*
4. Some other smaller grammatical changes are needed. For example on page 5 line 17, "differences of different" should be differences between". page 11, line 19, the statement "but different to several other studies" should be "different from other studies".	Thank you for pointing out this. Changes made to the text. Further, the article was proofread by an English-speaking colleague.
5. There are also some places in the article where repetition could be eliminated (such as the first paragraph under 'agreement and differences' on page 7).	Change made to the text. Line 209-211 "The Wilcoxon Signed-Ranks Test was used to assess the overall difference between the EQ-5D-3L and SF-6D utility scores, the difference of the utility scores according to different socio-demographic and disease related features."
6. In the abstract, the second-last sentence of the results section should ideally be re-worded to avoid any accidental implication that this is a longitudinal study. (Instead of saying that the scores decreased with advancing CKD stage, could say that they were significantly lower for those in more advanced stages.)	Change made to the text. Line 39-40 "Both MAUI scores were significantly lower ($p<0.001$;ANOVA) among those who were in more advanced stages of the disease"
7. When describing instruments' scoring, 1.00 on the EQ-5D is said to indicate 'good health status' (page 7, line 7). More accurately, it is meant to represent 'best imaginable health'. (The following sentence, stating that values close to zero indicate worse conditions while 1 represents perfect health, is repetitive and could be omitted.) To be consistent when describing the SF-6D, I suggest	Change made to the text. Line 186-189 "The EQ-5D utility scores range from -0.59, 0=being dead; negative values represent health status considered worse than "dead", to 1.00 which indicate best imaginable health. The SF-6D utility scores ranged from 0.296 which indicate

stating that its scores range from 0.296 to 1.0 (giving lowest value first).	severely impaired levels in all dimensions to 1.0 which indicates no difficulty in any dimensions."
8. Under 'Discrimination' on page 8, the first sentence (line 11) should state that MAUIs are meant to measure change in QOL (rather than just improvement), reflecting change in health among those with the condition of interest. It is good that at the end of page 11 the authors acknowledge that the different disease stages may not actually always represent different states of health-related quality of life. That is an important reason for measuring HRQOL, and could be mentioned in the other areas of the paper where discrimination is discussed.	Changes made to the text. Line 234
9. Our 2014 study seems relevant to cite: • Stewart, S. T., Cutler, D.M. & Rosen, A.B. (2014). Comparison of Trends in U.S. Health-Related Quality of Life over the 2000's Using the SF-6D, HALex, EQ-5D, and EQ-5D Visual Analog Scale versus a Broader Set of Symptoms and Impairments. Medical Care, 52(12), 1010-1016.	Cited the research paper.

VERSION 2 – REVIEW

REVIEWER	Shunping Li School of Health Care Management, Shandong University. China
REVIEW RETURNED	03-Oct-2018

GENERAL COMMENTS	The authors have addressed my concerns except for the statement of strengths of this study. I don't think that the high response rate and experienced data collectors should be the main strengths of this study, although these two statements is attracting. The main added information should be shown here, for instance, the SF-6D is more appropriate in CKD. Another minor misconception is the Page 6, line 51: SF-6D is derived from SF-36 V1, SF-36 V2 and SF-12 V2. Please check it again. Please uniform the format of SF-36, EQ-5D-3L and SF-6D. For example line 137, line 344.

REVIEWER	Susan Stewart National Bureau of Economic Research, U.S.A.
REVIEW RETURNED	10-Oct-2018

GENERAL COMMENTS	The authors have responded to my comments. Some small grammatical error remain, but presumably will be corrected in the

	final proof editing process. I make a few final comments and suggestions below. The authors have added points starting at line 52, under 'strengths and limitations of the study'. However it seems that two strengths that they deleted could be added back: This is the first study to compare the utility scores arising from the EQ-5D-3L and SF-6D in Chronic Kidney Disease patients and demonstrate that these tools cover different **aspects of** health among CKD patients. In the abstract under 'setting' and in the methods section, it would be informative to explain how the CKD patients were found; presumably through government or private medical records? At the top of page 6, not quite so much information is necessarily needed about the interviewers, but it should be clarified that interviews were done face-to-face (at home, or was it sometimes elsewhere?) line 170 states that the SF-6D is derived from SF-36, SF-12 Version 1 and SF-12 Version 2. It would be more clear to say that the SF-6D **can be** derived from either of these SF-12 questionnaires. line 193: grammar is unclear in this sentence" "Distribution of the socio-demographic characteristics of the study population was compared with their mean utility scores." Do you mean that "Mean utility scores on each instrument were compared by socio-demographic characteristics"? line 228 should state that MAUIs can measure both improvement and worsening of a condition: "MAUIs are meant to measure change in QOL due to health **change** in the condition of interest." On page 10, the paragraph should be more concise in describing results: "Further the two utility scores were significantly different among males (<0.001), those above age 40, those with education from grade 5 to General Certificate of Education (GCE) - Ordinary Level, those who were employed, those with and without comorbidities, and those up to stage IV of CKD and dialysis patients (Table 1)." (or, those with all stages of CKD except stage V") On the second half of page 10, it does not seem necessary to describe the correlation results in so much detail in the text since they are all in the table. line 331: "The test re-test ICC was more than 0.943 in EQ-5D-3L" (why 'more than', why not exactly 0.943?) line 352: I suggest adding the portion in bold to this sentence" "the responsiveness of the two instruments **to changes in kidney function over time** was not assessed." Also, this fact seems to belong better in the limitations section? end of page 11: this fact also seems to belong better in the limitations section? I also make suggested changes to these sentences. It is best to differentiate between utility valuations/values

	and scores (here and in lines 160 and 162): "Though Sri Lankan EQ-5D-3L utility values scores are available (18), yet we used the UK utility values scores for the EQ-5D-3L (33) because of the unavailability of comparable Sri Lankan SF-6D utility scores values. The conclusion seems to say twice that the SF-6D was better at detecting different stages of the disease.

VERSION 2 – AUTHOR RESPONSE

Comment	**Response**
Reviewer: 1	
Another minor misconception is the Page 6, line 51: SF-6D is derived from SF-36 V1, SF-36 V2 and SF-12 V2. Please check it again.	According to the literature SF 6D can be derived both form SF 36 and SF 12 (V1 & V2) *Busija, L., Pausenberger, E., Haines, T.P., Haymes, S., Buchbinder, R. and Osborne, R.H., 2011. Adult measures of general health and health-related quality of life: Medical Outcomes Study Short Form 36-Item (SF-36) and Short Form 12-Item (SF-12) Health Surveys, Nottingham Health Profile (NHP), Sickness Impact Profile (SIP), Medical Outcomes Study Short Form 6D (SF-6D), Health Utilities Index Mark 3 (HUI3), Quality of Well-Being Scale (QWB), and Assessment of Quality of Life (AQOL). Arthritis care & research, 63(S11), pp.S383-S412.*
Please uniform the format of SF-36, EQ-5D-3L and SF-6D. For example, line 137, line 344.	Changes made
Reviewer: 2	
The authors have added points starting at line 52, under 'strengths and limitations of the study'. However it seems that two strengths that they deleted could be added back	Will not attend to this as instructed by the editor
In the abstract under 'setting' and in the methods section, it would be informative to explain how the CKD patients were found; presumably through government or private	Following was added to the 'Setting' section of the abstract (Line 31).

medical records.	"A representative sample of 1096 CKD patients, selected using the population-based CKD register, completed the EQ-5D-3L and SF-36".
At the top of page 6, not quite so much information is necessarily needed about the interviewers, but it should be clarified that interviews were done face-to-face (at home, or was it sometimes elsewhere?)	Changes made (Line 146-147)
Line 170 states that the SF-6D is derived from SF-36, SF-12 Version 1 and SF-12 Version 2. It would be more clear to say that the SF-6D can be derived from either of these SF-12 questionnaires.	Changes made (Line 171) "SF-6D is derived from either SF-36 or SF-12 (Version 1 and Version 2)"
Line 193: grammar is unclear in this sentence" "Distribution of the socio-demographic characteristics of the study population was compared with their mean utility scores." Do you mean that "Mean utility scores on each instrument were compared by socio-demographic characteristics"?	Changes made (Line 194) "Mean utility scores on each instrument were compared by socio-demographic characteristics"
Line 228 should state that MAUIs can measure both improvement and worsening of a condition: "MAUIs are meant to measure change in QOL due to health change in the condition of interest."	Changes made (Line 231) "It is important that MAUIs can discriminate correctly among groups of different severity as MAUIs are meant to measure change in QOL due to improvement or worsening of the health in the condition of interest".
On page 10, the paragraph should be more concise in describing results: "Further the two utility scores were significantly different among males (<0.001), those above age 40, those with education from grade 5 to General Certificate of Education (GCE) - Ordinary Level, those who were employed, those with and without comorbidities, and those up to stage IV of CKD and dialysis patients (Table 1)." (or, those with all stages of CKD except stage V")	Changes made (Line 291)
On the second half of page 10, it does not seem necessary to describe the correlation results in	Changes made (Line 308)

so much detail in the text since they are all in the table.	
Line 331: "The test re-test ICC was more than 0.943 in EQ-5D-3L" (why 'more than', why not exactly 0.943?)	Changes made (Line 334)
Line 352: I suggest adding the portion in bold to this sentence" "the responsiveness of the two instruments to changes in kidney function over time was not assessed." Also, this fact seems to belong better in the limitations section?	Changes made (Line 355)
End of page 11: this fact also seems to belong better in the limitations section? I also make suggested changes to these sentences. It is best to differentiate between utility valuations/values and scores (here and in lines 160 and 162): "Though Sri Lankan EQ-5D-3L utility values scores are available (18), yet we used the UK utility values scores for the EQ-5D-3L (33) because of the unavailability of comparable Sri Lankan SF-6D utility scores values.	Thank you for the suggestion. However, the author believe this is best suited in the discussion section. Suggested changes were done to the sentence structure.
The conclusion seems to say twice that the SF-6D was better at detecting different stages of the disease.	Changes made (Line 410)

VERSION 3 – REVIEW

REVIEWER	Shunping Li Shandong University, China
REVIEW RETURNED	10-Dec-2018

GENERAL COMMENTS	The authors have addressed all my concerns.

REVIEWER	Susan Stewart National Bureau of Economic Research 1050 Massachusetts Ave Cambridge, MA 02138
REVIEW RETURNED	05-Dec-2018

GENERAL COMMENTS	A few small grammatical issues may remain but I trust that they will be picked up during the proof process if problematic. For example on line 227 "improvement or worsening of the health," should be 'of health (no comma)') On lines 286-287, this phrase: "more than 40 years groups, those who were employed, among both who had and didn't have comorbidities" might read more clearly as: "age groups above age 40, those who were employed, and those with and without comorbidities" On lines 343-344, the phrase "though Sri Lankan EQ-5D-3L utility scores are available (18), yet we used the UK utility scores" should omit the word 'yet' I suggest that this sentence from the conclusion be changed to past tense to match the others (lines 396-97), change 'is' to 'was': "Effect size, which denoted the discriminating ability of the different CKD stages, is highest when disease condition is advanced."