PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Diagnostic accuracy of self-administered urine glucose test strips as a diabetes screening tool in a low-resource setting in Cambodia
AUTHORS	Storey, Helen; van Pelt, Maurits; Bun, Socheath; Daily, Frances; Neogi, Tina; Thompson, Matthew; McGuire, Helen; Weigl, Bernhard

VERSION 1 – REVIEW

REVIEWER	Ann Van den Bruel Nuffield Department of Primary Care Health Sciences, University of Oxford, UK
REVIEW RETURNED	22-Oct-2017

GENERAL COMMENTS	The study presents interesting findings on the accuracy of urine test strips for diabetes in the general population. Although I think the study is overall well conducted, I do have some comments which I would like the authors to address. 1. The validity of the references standard: by their own admission, the reference standard which was used here is not the gold standard. Although limitations of the OGTT were highlighted, one third of diabetes patients were diagnosed based on their capillary fasting blood glucose levels. How reliable is this two-step approach? 2. Capillary fasting blood glucose is used as a reference standard and an index test. Although different thresholds were used, patients with capillary fasting blood glucose of at least 200 mg/dl did not proceed to OGTT and were diagnosed using the capillary fasting blood glucose only. This leads to incorporation bias, inflating the accuracy of the test. In that respect, I found the addition of the two other index test distracting from the central message which is on urine test strips. 3. Differences between subgroups: no differences were found based on overlapping confidence intervals. Although this is common practice, it might miss true differences because by dividing the population into subgroups you decrease precision and confidence intervals will be wider. It would have been more appropriate to calculate diagnostic accuracy of the index test in combination with clinical features in a logistic regression analysis. 4. The authors discuss the possibility of selecting a more at risk population in whom the urine tests strips might be more efficacious. This is a very interesting hypothesis, which could be tested with the data reported in this study. Based on the logistic regression analysis in my previous point, the authors could analyse accuracy of urine test strips in a pre-selected population based on clinical features. 5. The abstract is full with abbreviations that only become clear after reading the manuscript. I would advise the authors to rewrite these to improve readability.

REVIEWER	Jan Verbakel Nuffield Department of Primary Care Health Sciences, University of Oxford
REVIEW RETURNED	07-Nov-2017

GENERAL COMMENTS	Dear Editor, Thank you for this opportunity to review the paper by Storey et al., entitled: Diagnostic accuracy of self-administered urine glucose test strips as a diabetes screening tool in a low-resource setting in Cambodia. The authors have studied a very relevant clinical problem and have been able to assess the diagnostic accuracy of self-administered urine glucose test strips to detect diabetes in a low-resource country. There are a few minor issues, which I believe need to be addressed. MINOR ISSUES: ---------- Abstract: ---------- - Outcomes: UGTS abbreviations has not been explained previously. Pleas revise. As the cFBG (index test) is also part of the composite reference standard, there might have been some incorporation bias, although assessed at a higher threshold in the reference test. This should commented on in the discussion section. Methods: ---------- - is there a protocol available for this study? And if so, were the subgroup analyses prespecified? If not, performing significance testing on these subgroups might seem a bit arbitrary. - line 103: oral or written informed consent? - line 115: any previous data available on the analytical accuracy of the On Call Plus glucometer? Would be relevant to state here. - line 154: "continuous values were compared using Student's t-test": Maybe useful to mention what was compared exactly, e.g. mean value. Did you examine whether these variables were normally distributed. If not, the t-test would not be appropriate. Results: ---------- - line 162: "reported taking medication for diabetes that day": shouldn't they be excluded from the recruitment considering they should have been off diabetes medication for 30 days? - line 164: similar to previous comment: these 8% should not have been included in the first place? - line 168: please provide confidence intervals for your estimates of sens/spec. - Table 3 & 4: Considering the imbalance in sample size per subgroup (7 versus 1048), using p-values to assess statistical significance seems a bit exaggerated. I can imagine these analyses were decided upon post-hoc, so I would suggest to keep these as descriptive as possible. Discussion: ----------

	- line 193: I agree that continuous high glucose levels contributes to higher risk of long-term complications (as per reference 20), but does a single high glucose measurement at screening have the same long-term effect? - line 232: It would be good to provide a reference here concerning the impact of anaemia on test performance.

REVIEWER	Elizabeth Dehmer, MD, MPH PhD candidate UNC Gillings School of Global Public Health UNC Chapel Hill Chapel Hill, NC USA Board certified nephrologist Southeastern Nephrology Associates Wilmington, NC USA
REVIEW RETURNED	01-Dec-2017

GENERAL COMMENTS	This is a well-written paper which follows the STARD guidelines for a study of diagnostic accuracy. The discussion of study limitations was very thorough. The conclusion was appropriate based on study results.

VERSION 1 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:

Reviewer: 1
Reviewer Name: Ann Van den Bruel
Institution and Country: Nuffield Department of Primary Care Health Sciences, University of Oxford, UK Please state any competing interests or state 'None declared': I hereby state that I have no competing interests.
Please leave your comments for the authors below.

The study presents interesting findings on the accuracy of urine test strips for diabetes in the general population.
Although I think the study is overall well conducted, I do have some comments which I would like the authors to address.
1. The validity of the references standard: by their own admission, the reference standard which was used here is not the gold standard. Although limitations of the OGTT were highlighted, one third of diabetes patients were diagnosed based on their capillary fasting blood glucose levels. How reliable is this two-step approach?
The need for this two-step approach was based on what was considered clinically safe and appropriate for the patients. Of the 70 people who had a cFBG>200, all also had a A1c value greater than 6.5%, which is another marker of diabetes. Also, there was one person who had a cFBG>200 who inadvertently took the OGTT, and that person was also positive by OGTT. Based on this data, we feel the two-step approach is reliable and we have now noted this in paragraph 2 of the results.

Additionally, when reviewing the data to answer this comment, we noted that there was a mistake in the number of people who were cFBG positive (70 rather than 74). This mistake has now been corrected in the manuscript.

2. Capillary fasting blood glucose is used as a reference standard and an index test. Although different thresholds were used, patients with capillary fasting blood glucose of at least 200 mg/dl did not proceed to OGTT and were diagnosed using the capillary fasting blood glucose only. This leads to incorporation bias, inflating the accuracy of the test. In that respect, I found the addition of the two other index test distracting from the central message which is on urine test strips.
This is an important point. We included comparison of the other 2 index tests as these are likely alternatives to UGTS as a screening test. In paragraph 6 of the discussion we discuss some pros and cons of each test, as they relate to use in our setting. However, the incorporation bias is an important point, so we have now commented on this in the discussion in paragraph 4.

3. Differences between subgroups: no differences were found based on overlapping confidence intervals. Although this is common practice, it might miss true differences because by dividing the population into subgroups you decrease precision and confidence intervals will be wider. It would have been more appropriate to calculate diagnostic accuracy of the index test in combination with clinical features in a logistic regression analysis.
Thank you for this comment as we had not investigated it prior. We evaluated univariate and multivariate logistic regression models including age, BMI, gender, and waist circumference as independent clinical variables in the model. None of the clinical features were statistically significant so we are more confident now that we did not miss a true difference. We have also added a sentence describing this to the reader in paragraph 3 of the results.

4. The authors discuss the possibility of selecting a more at-risk population in whom the urine tests strips might be more efficacious. This is a very interesting hypothesis, which could be tested with the data reported in this study. Based on the logistic regression analysis in my previous point, the authors could analyse accuracy of urine test strips in a pre-selected population based on clinical features.
Based on our results from the logistic regression analysis above, the clinical features we examined may not have an impact on the accuracy of the urine glucose test strip. However, many other clinical features remain and a more exhaustive study of this may still be useful in future studies. We also added a note that in this study, the sensitivity of the urine glucose test strip doubled among overweight men with a high waist circumference (29% vs. 14% respectively). This text is in paragraph 7 of the discussion.

5. The abstract is full with abbreviations that only become clear after reading the manuscript. I would advise the authors to rewrite these to improve readability.
We have now spelled out the abbreviations in the abstract to be clearer to the reader.

Reviewer: 2
Reviewer Name: Jan Verbakel
Institution and Country: Nuffield Department of Primary Care Health Sciences, University of Oxford
Please state any competing interests or state 'None declared': None declared
Please leave your comments for the authors below

Thank you for this opportunity to review the paper by Storey et al., entitled: Diagnostic accuracy of self-administered urine glucose test strips as a diabetes screening tool in a low-resource setting in Cambodia. The authors have studied a very relevant clinical problem and have been able to assess the diagnostic accuracy of self-administered urine glucose test strips to detect diabetes in a low-resource country.
There are a few minor issues, which I believe need to be addressed.

MINOR ISSUES:

Abstract:

- Outcomes: UGTS abbreviations has not been explained previously. Pleas revise.
Abbreviations in the abstract have now been spelled out.

As the cFBG (index test) is also part of the composite reference standard, there might have been some incorporation bias, although assessed at a higher threshold in the reference test. This should commented on in the discussion section.
The incorporation bias is an important point, also noted by reviewer 1, so we have now commented on this in the discussion in paragraph 4.

Methods:

- is there a protocol available for this study? And if so, were the subgroup analyses prespecified? If not, performing significance testing on these subgroups might seem a bit arbitrary.
The subgroup analyses were not prespecified, and we note this in the methods. We agree that as exploratory analyses, the purpose is more for hypothesis generation rather than confirmation. And though arbitrary, they may still be informative to the reader when considering the results.
- line 103: oral or written informed consent?
Consent was written and we have now noted that in the methods.
- line 115: any previous data available on the analytical accuracy of the On Call Plus glucometer? Would be relevant to state here.
The 510(k) substantial equivalence determination decision summary is available from the FDA for those who are interested: https://www.accessdata.fda.gov/cdrh_docs/reviews/K090057.pdf. We have added mention of the product website in paragraph 2 of the methods for those readers who would like more information on the product specifications.
- line 154: "continuous values were compared using Student's t-test": Maybe useful to mention what was compared exactly, e.g. mean value. Did you examine whether these variables were normally distributed. If not, the t-test would not be appropriate.
Yes, mean values were compared after confirming the normality of the data. We note this in the tables but not the methods, so we have now added it to the methods for completeness.

Results:

- line 162: "reported taking medication for diabetes that day": shouldn't they be excluded from the recruitment considering they should have been off diabetes medication for 30 days?
Yes, we agree they should have been excluded from the study if they had taken diabetes medication in the last 30 days. There were two points of contact with the patient, once at screening where the UGTS was provided for self-administration, and once at follow up where patients were clinically assessed. The 6 patients that were excluded from the analysis for taking diabetes medication may have answered differently at the two visits, which is why it was important to ask the question twice.
- line 164: similar to previous comment: these 8% should not have been included in the first place?
Similar to our response above, it is possible that the participant answered the questions at screening and follow up inconsistently. As a further check to our analysis, we confirmed that this clinical feature did not statistically impact the accuracy of the urine glucose test strip using logistic regression analyses.
- line 168: please provide confidence intervals for your estimates of sens/spec.
We now have the confidence intervals in the text as they are in table 2.
- Table 3 & 4: Considering the imbalance in sample size per subgroup (7 versus 1048), using p-values to assess statistical significance seems a bit exaggerated. I can imagine these analyses were decided upon post-hoc, so I would suggest to keep these as descriptive as possible.

Similar to our response above, the subgroup analyses were not prespecified, and therefore are exploratory in nature. Though the sample size is imbalanced, it does not invalidate the analysis, and we share this with the readers for their better understanding of the data.

Discussion:

- line 193: I agree that continuous high glucose levels contributes to higher risk of long-term complications (as per reference 20), but does a single high glucose measurement at screening have the same long-term effect?
We agree with your comment as our study is cross-sectional in nature and therefore, cannot speak to longer effects. The sentence has now been rephrased to emphasize our uncertainty about long term effects.
- line 232: It would be good to provide a reference here concerning the impact of anaemia on test performance.
We have now added a reference to this sentence for further consideration by the reader.

Reviewer: 3
Reviewer Name: Elizabeth Dehmer, MD, MPH Institution and Country: PhD candidate, UNC Gillings School of Global Public Health, UNC Chapel Hill, Chapel Hill, NC, USA and Board certified nephrologist, Southeastern Nephrology Associates, Wilmington, NC, USA Please state any competing interests or state 'None declared': None declared
Please leave your comments for the authors below

This is a well-written paper which follows the STARD guidelines for a study of diagnostic accuracy. The discussion of study limitations was very thorough. The conclusion was appropriate based on study results.
Thank you for the supportive comments.

FORMATTING AMENDMENTS (if any)
Required amendments will be listed here; please include these changes in your revised version:
We noted some discrepancies in the footnoting of our tables and we have now cleaned that up for better consistency.

VERSION 2 – REVIEW

REVIEWER	Ann Van den Bruel University of Oxford, UK
REVIEW RETURNED	15-Jan-2018

GENERAL COMMENTS	The revisions have greatly improved the paper. I have only one, minor comment: could you please add a statement to the discussion on the risk of incorporation bias from using capillary glucose measurements for the reference standard and as index test

REVIEWER	Elizabeth Dehmer, MD, MPH PhD candidate, UNC Gillings School of Global Public Health, Chapel Hill, NC, USA
REVIEW RETURNED	26-Jan-2018

GENERAL COMMENTS	I think the authors responded appropriately to reviewer concerns. There remains some inconsistency with the abbreviations - although spelled out in the abstract, there are abbreviations in the Strengths and Limitations section (OGTT for example) that are not previously abbreviated in the text. Also, urine glucose test strip is spelled out except on pg 12 abbreviated as UGTS. In the background section, there are a few abbreviations (International Diabetes Federation IDF and non-communicable diseases NCDs) which I do not think are used again and so could be omitted. I'd err on the side of less abbreviations.

REVIEWER	Jan Verbakel Nuffield Department of Primary Care Health Sciences, University of Oxford Department of Public Health and Primary Care, KULeuven
REVIEW RETURNED	29-Jan-2018

GENERAL COMMENTS	The authors have addressed all of my comments appropriately. Just one minor issue: I would like to suggest to remove the p-values from page 9, line 184-188 as these are not very informative (and biased by post-hoc subgroup analyses) and even not mentioned in the tables, but I leave this up to the editor to decide.

VERSION 2 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:

Reviewer: 1
Reviewer Name: Ann Van den Bruel
Institution and Country: University of Oxford, UK
Please state any competing interests or state 'None declared': None declared
Please leave your comments for the authors below

The revisions have greatly improved the paper.
I have only one, minor comment: could you please add a statement to the discussion on the risk of incorporation bias from using capillary glucose measurements for the reference standard and as index test.
Yes, we had a brief statement on this in paragraph 4 of the discussion and have now added more explanation. We agree it is an important consideration.

Reviewer: 3
Reviewer Name: Elizabeth Dehmer, MD, MPH
Institution and Country: UNC Gillings School of Global Public Health, Chapel Hill, NC, USA
Please state any competing interests or state 'None declared': None declared
Please leave your comments for the authors below

I think the authors responded appropriately to reviewer concerns.

There remains some inconsistency with the abbreviations - although spelled out in the abstract, there are abbreviations in the Strengths and Limitations section (OGTT for example) that are not previously abbreviated in the text. Also, urine glucose test strip is spelled out except on pg 12 abbreviated as

UGTS. In the background section, there are a few abbreviations (International Diabetes Federation IDF and non-communicable diseases NCDs) which I do not think are used again and so could be omitted. I'd err on the side of less abbreviations.
Thank you for this attention to detail. We have spelled out OGTT in the strengths/limitations section, UGTS on page 12, and IDF, NCD, CVD, and WHO in the background. They are also removed from the list of abbreviations at the end. We agree consistency with abbreviations is helpful for the reader.

Reviewer: 2
Reviewer Name: Jan Verbakel
Institution and Country: 1. Nuffield Department of Primary Care Health Sciences, University of Oxford 2. Department of Public Health and Primary Care, KULeuven
Please state any competing interests or state 'None declared': None declared.
Please leave your comments for the authors below

The authors have addressed all of my comments appropriately. Just one minor issue: I would like to suggest to remove the p-values from page 9, line 184-188 as these are not very informative (and biased by post-hoc subgroup analyses) and even not mentioned in the tables, but I leave this up to the editor to decide.
We have elected to remove the p-values from this text. Thank you for the suggestion.