PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Diagnostic accuracy of the Xpert CT/NG and OSOM Trichomonas Rapid assays for point-of-care STI testing among young women in South Africa: a cross-sectional study
AUTHORS	Garrett, Nigel; Mitchev, Nireshni; Osman, Farzana; Naidoo, Jessica; Dorward, Jienchi; Singh, Ravesh; Ngobese, Hope; Rompalo, Anne; Mlisana, Koleka; Mindel, Adrian

VERSION 1 – REVIEW

REVIEWER	Remco Peters Department of Medical Microbiology University of Pretoria, South Africa
REVIEW RETURNED	15-Oct-2018

GENERAL COMMENTS	The manuscript by Nigel Garrett and colleagues is well-written and presents relevant data on the diagnostic accuracy of Xpert CT/NG and OSOM TV assays for STI diagnostics and treatment in clinical settings in South Africa. The data are relevant to support the debate around replacing syndromic by aetiological management of STIs in low- and middle-income countries. Please find some comments and suggestions below, in particular with regard to positioning of the OSOM TV assay in the manuscript and including some data on microbial load. The evaluation of the OSOM TV Assay is not reported consistently in the manuscript. For example, the assay is not mentioned in the Introduction. The manuscript is positioned around utilising the existing Xpert platform potential to introduce STI testing. However, the authors use a different TV test and did not include Xpert TV testing. My suggestion would be to strengthen positioning of the TV assay evaluation throughout the manuscript, or to delete it from the paper. Abstract: the evaluation of diagnostic performance of the TV assay is not presented consistently; reference to it should be included in the introduction and conclusion sections. Strength and limitations, first bullet: this is not the first evaluation of Xpert in a LMIC, but the first evaluation of diagnostic performance of this assay in a LMIC. Methods: There has been a lot of debate about Xpert for TB diagnostics are PHC facilities. The consensus is that this is considered a near-patient test, but does not meet the criteria to be called a point of care test. I don't think it is appropriate to call Xpert CT/NG a POC test.

	Methods: what treatment was provided to participating women? Targeted following Xpert result or syndromic? Methods: what is the turnaround time for the OSOM TV test? Were these results used to determine treatment of participants? Introduction/Methods: was the OSOM TV test ever evaluated before in other settings? What was the analytical and diagnostic performance? Why was this test chosen in this evaluation instead of e.g. Xpert TV Assay and/or wet mount microscopy as POC test? Results, first paragraph: the manuscript is about diagnostic performance which means the results of two tests (swabs) should be available; your sample size is 247. This means the data in the first sentences of this paragraph should be reported on 247 instead of 267 patients. Results: it would be good to report on cycle threshold values for discordant specimens where Xpert is positive but STI-7 negative. Are these all low load specimens, i.e. sampling variation, or are there also discordant strong-positive specimens? Results: same for the one specimen that is Xpert NG positive only: did both probes react? And to what fluorescence level/how far above compared to the cut-off value? Results: most molecular assays suggest to repeat low-positive reactions (e.g. cycle threshold value above 37 or 38 cycles). Is this the same for all the assays evaluated in this manuscript? Or could that explain discordant results? Results: I know it is not the objective of the paper, but the reader would be curious what the detection rates were for the other pathogens included in the STI-7 assay… perhaps it would be valuable to report these in a single or two sentences? Discussion: "(..)limited by a moderate sensitivity, as shown in previous studies". Please provide reference and range reported in literature. Discussion: the authors should discuss/interpret the discordant specimens: are these true positives detected by Xpert? Or very low load specimens that are potentially false-positives, cross reaction (e.g. with other Neisseria species) or local contamination? Discussion: another limitation is that specimens were not available to repeat discordant specimens on the Xpert CT/NG platform.

REVIEWER	Louise Causer The Kirby Institute, UNSW Sydney
REVIEW RETURNED	29-Oct-2018

GENERAL COMMENTS	This is an interesting brief report of diagnostic accuracy of POC testing for CT/NG and TV from Durban, South Africa. As noted, it is unique for a performance evaluation of Xpert CT/NG performance in being from LMIC country. Some suggestions below for consideration by the authors: In the Abstract:

	Objectives - the aim specifies only to evaluate the performance of the Xpert CT/NG. There is no mention of OSOM POC test for TV until the methods. Please revise to include. Methods - expand FTD STD9 in the abstract. Please also include study dates in abstract. Conclusion - again, no mention of TV POC results here. Please revise to be consistent with methods. Overall: Consider including a definition of POC test as there are different opinions in the literature about Xpert being a POC test vs a near-patient tests. Be consistent in use of POC test and POC test result. Introduction - Please revise to include information related to TV testing and the POC test for TV evaluated here. Again, only Xpert CT/NG is included. Given that Treponema pallidum is mentioned in the opening paragraph, authors might consider a brief mention of how syphilis is detected in this setting. Are POC tests also used for syphilis at all here? The authors note the potential existing infrastructure available as a result of TB testing with Xpert. How practically available would these same devices be for doing STI tests? Methods - please include exclusion criteria here. Please add some more detail and references to support use of the comparison assays Anyplex II and the FTD STD9. Please state why these assays were chosen and considered as the gold-standard for these analyses? Please add some brief details on how Xpert CT/NG and OSOM were performed. Or provide reference to manufacturer intructions. These tests operate quite differently and readers may like to have some brief information on how they work (if word count allows). Data analysis - Please include information related to sample size calculation for this study. Patient and Public Involvement - The section needs further work, perhaps elaboration of the statements in first two sentences as currently adds little to the paper overall. Results - please add flow chart as per STARD guidelines. Is there any information to explain the 2 invalid results? Is there any information on the crossing point/thresholds for the discordant CT or NG tests positive on the Xpert? Previous work referenced (Causer et al) suggested these discordants were perhaps related to low organism loads, at threshold of assay detection. It would be interesting to include this information from your study if available. What was the time delay between specimen collection and references laboratory testing? How were specimens transported etc? Could this have impacted on specimen quality, contributing to the CT and NG discordants observed? Discussion - the first sentence includes evaluation of performance of both Xpert CT/NG and OSOM TV - however as noted earlier, this is not stated in the intro or abstract objectives. Please add accordingly. The CAPRISA protocol included with this submission mentions the Xpert CT/NG and TV assays, not the OSOM TV test evaluated here. Why was the Xpert TV assay not evaluated? This would seem like the most appropriate POC tests to use in this setting? Understand it perhaps was not available - however this should be noted in the discussion and/or intro as it is an obvious option as is now available.

	Final paragraph should address TV POC test also as this was included in the study reported here.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1: Prof Remco Peters	**Authors' Reply**
The manuscript by Nigel Garrett and colleagues is well-written and presents relevant data on the diagnostic accuracy of Xpert CT/NG and OSOM TV assays for STI diagnostics and treatment in clinical settings in South Africa. The data are relevant to support the debate around replacing syndromic by aetiological management of STIs in low- and middle-income countries. Please find some comments and suggestions below, in particular with regard to positioning of the OSOM TV assay in the manuscript and including some data on microbial load.	We thank Prof Peters for his insightful comments and for recognizing the importance of our work. We have strengthened the evaluation of the OSOM TV assay throughout the paper, and, as suggested, we have also added more information about the discordant STI results including PCR cycle threshold as an indicator of microbial load.
The evaluation of the OSOM TV Assay is not reported consistently in the manuscript. For example, the assay is not mentioned in the Introduction. The manuscript is positioned around utilising the existing Xpert platform potential to introduce STI testing. However, the authors use a different TV test and did not include Xpert TV testing. My suggestion would be to strengthen positioning of the TV assay evaluation throughout the manuscript, or to delete it from the paper.	As suggested, we have strengthened the information and discussion provided about the OSOM TV assay throughout the manuscript (please see reply to individual comments below).
Abstract: the evaluation of diagnostic performance of the TV assay is not presented consistently; reference to it should be included in the introduction and conclusion sections.	We have added OSOM TV test to the abstract and the following paragraph to the Introduction: '***Considering the relatively high cost of the Xpert Trichomonas vaginalis (TV) cartridge (~ USD 19.00), we decided to complement the Xpert CT/NG with the OSOM TV antigen detection assay (Sekisui, Lexington, MA, US) and Gram stain microscopy, in order to offer the participants a comprehensive 2-hour STI testing alternative to syndromic management (10). The advantage of the OSOM TV assay is that it is relatively cheap (~ USD 8.00), has a rapid processing time (~ 10 minutes), and has shown higher accuracy than wet mount microscopy, especially in women (11, 12).'*** Additional text was added to

	the Discussion section – please see comments below.
Strength and limitations, first bullet: this is not the first evaluation of Xpert in a LMIC, but the first evaluation of diagnostic performance of this assay in a LMIC.	We have amended the sentence as suggested: *'This is **the first evaluation of the diagnostic performance** of the Xpert CT/NG point-of-care (POC) assay to detect chlamydia and gonorrhoea from a low- and middle-income country'*
Methods: There has been a lot of debate about Xpert for TB diagnostics at PHC facilities. The consensus is that this is considered a **near-patient test**, but does not meet the criteria to be called a point of care test. I don't think it is appropriate to call Xpert CT/NG a POC test.	In response to both Reviewers' comments we have added the following sentence and POC definition to the Methods section: ***'Considering that all participants received their results during the same visit and the test were performed in the clinic, we used the term POC for both assays, in line with the following consensus definition: a 'point-of-care test…is a test to support clinical decision making, which is performed by a qualified…staff nearby the patient…during or very close to the time of consultation, to help the patient and physician to decide upon the best suited approach, and of which the results should be known at the time of the clinical decision making' (15).'***
Methods: what **treatment** was provided to participating women? Targeted following Xpert result or syndromic?	We have added information on the treatment provided to the participants in the study, which is described in more detail in Reference 9: ***'Women diagnosed with CT, NG or TV based on POC testing were offered immediate supervised treatment with single dose antibiotics on the same visit. Treatment regimens followed international guidelines, and were compatible with national guidelines: Ceftriaxone 250mg intramuscular and Azithromycin 1g oral for NG, Azithromycin 1g oral for CT, and Metronidazole 2g oral for TV (13, 14).'***
Methods: what is the turnaround time for the OSOM TV test? Were these results used to determine treatment of participants?	We added the turnaround time '***rapid processing time (~ 10 minutes)'*** to the Introduction and clarified that the participants were treated '***based on POC testing'*** in the Methods section.
Introduction/Methods: was the OSOM TV test ever evaluated before in other settings? What was the analytical and diagnostic performance? Why was this test chosen in this evaluation	The OSOM TV was previously evaluated in other settings and we have added additional details on diagnostic performance to the Introduction and Discussion sections: …***'has shown higher accuracy than wet mount microscopy, especially in women (11, 12).'***

instead of e.g. Xpert TV Assay and/or wet mount microscopy as POC test?	and ***'Previous evaluations (11, 12) of the OSOM TV concur with our finding of a slightly lower sensitivity (92 – 98%) than PCR technology, but a consistently high specificity to detect TV (99%).'*** We have also added an explanation to the Introduction section that we chose the OSOM TV over the Xpert TV assay based on cost (see comment reply above).
Results, first paragraph: the manuscript is about diagnostic performance which means the results of two tests (swabs) should be available; your sample size is 247. This means the data in the first sentences of this paragraph should be reported on 247 instead of 267 patients.	We have clarified the distinction between the number of participants enrolled into the ***'CAPRISA 083'*** cohort study and the number included, '247/267 ***(92.5%)***', in the present diagnostic evaluation. We felt it was important to show the reader that most participants from the CAPRISA 083 study entered the diagnostic evaluation, thereby limiting potential selection bias. As suggested by Dr Causer, to further clarify this point, we have added a study flow diagram to the manuscript (see **Figure 1**).
Results: it would be good to report on **cycle threshold values for discordant specimens where Xpert is positive but STI-7 negative**. Are these all low load specimens, i.e. sampling variation, or are there also discordant strong-positive specimens?	Thank you for this suggestion. We have added this information to the Results section of the manuscript: ***'The Xpert cycle thresholds of the five discordant results reached 26.3 to 38.6 cycles, with two values greater than 38 cycles, indicating potential sampling or testing variation.'***
Results: same for the one specimen that is Xpert NG positive only: did both probes react? And to what fluorescence level/how far above compared to **the cut-off value**?	In response to the Prof Peters' comments, we re-checked all assay results in the study. It was noted that the Xpert results for one participant were incorrectly entered into the Redcap database despite two quality control checks. Both results (CT and NG) were reported as negative on the Xpert report, but entered as positive in the database. This error has been rectified and we have adjusted the sensitivity and specificity analysis in Table 1. This also means that there were only 5 (not 6) 'false positive' CT results and no (rather than 1) discordant NG result in the study. We have updated the Result section accordingly.
Results: most molecular assays suggest to repeat low-positive reactions (e.g. cycle threshold value **above 37 or 38 cycles**). Is this the same for all the assays evaluated in this manuscript? Or could that explain discordant results?	The two results with cycle threshold greater 37 or 38 cycles were not repeated in our study. We have added your suggested sentence to the limitation section of the discussion (see comment below).

Results: I know it is not the objective of the paper, but the reader would be curious what the detection rates were for the **other pathogens included in the STI-7 assay**… perhaps it would be valuable to report these in a single or two sentences?	We have added the following sentence to the Results section: ***'In addition, Anyplex testing revealed a 4.9% (2.2-7.5) prevalence of Mycoplasma genitalium, 33.6% (27.7-39.5) Mycoplasma hominis, 51.8% Ureaplasma parvum and 19.0% (14.1-23.9) Ureaplasma urealyticum.'***
Discussion: "(..)limited by a moderate sensitivity, as shown in previous studies". Please provide reference and range reported in literature.	We have revised the Discussion section and added the following sentence: ***'Previous evaluations (11, 12) of the OSOM TV concur with our finding of a slightly lower sensitivity (92 – 98%) than PCR technology, but a consistently high specificity to detect TV (99%).'***
Discussion: the authors should discuss/interpret the discordant specimens: are these true positives detected by Xpert? Or very low load specimens that are potentially false-positives, cross reaction (e.g. with other Neisseria species) or local contamination?	We have added the following paragraph to the Discussion section: ***'In our study, two out of five 'false positive' Xpert CT results had a high cycle threshold above 38 cycles. A validation of 50 randomly selected vulvovaginal samples from a study of pregnant women in Pretoria, South Africa (16) also found two positive Xpert CT/NG results with high cycle thresholds that were confirmed negative after testing on the PrestoPlus CT/NG/TV (Microbiome, Ltd., Houten, The Netherlands) and Anyplex assays. This highlights that some caution is needed when interpreting high cycle threshold results, and that retesting may be considered.'*** Please also note that there were no discordant NG results after the corrections described above.
Discussion: another limitation is that specimens were not available to repeat discordant specimens on the Xpert CT/NG platform.	As suggested, this sentence was added to the limitations section: ***'A further limitation was that specimens were not available to repeat discordant specimens on the Xpert CT/NG platform.'***
Reviewer 2: Dr Louise Causer	**Authors' Reply**
This is an interesting brief report of diagnostic accuracy of POC testing for CT/NG and TV from Durban, South Africa. As noted, it is unique for a performance evaluation of Xpert CT/NG performance in being from LMIC country. Some suggestions below for consideration by the authors:	We thank Dr Causer for her review of the manuscript and for recognizing the contribution our work makes to the field. We have addressed her comments below. Some comments were also raised by Prof Peters, and were addressed above.

Abstract	
Objectives - the aim specifies only to evaluate the performance of the Xpert CT/NG. There is no mention of OSOM POC test for TV until the methods. Please revise to include.	As suggested, we have added the OSOM TV assay evaluation to the objective in the abstract: '*We aimed to evaluate the diagnostic performance of the Xpert Chlamydia trachomatis/Neisseria gonorrhoeae (CT/NG)* ***and OSOM Trichomonas (TV) Test*** *as part of a STI care model for young women in South Africa.*'
Methods - expand FTD STD9 in the abstract. Please also include study dates in abstract.	We have included these changes in the abstract.
Conclusion - again, no mention of TV POC results here. Please revise to be consistent with methods.	We have amended the Conclusion of the abstract, and have added additional information about the OSOM TV in the Introduction and Discussion sections of the main manuscript. '*The Xpert CT/NG showed high accuracy among young South African women* ***and combined with the OSOM TV proved a useful tool*** *in this high HIV/STI burden setting.*'
Overall: Consider including a definition of POC test as there are different opinions in the **literature about Xpert being a POC test vs a near-patient tests.** Be consistent in use of POC test and POC test result.	As suggested, we have added one consensus definition of a POC test to the Methods section of the manuscript: '***Considering that all participants received their results during the same visit and the test were performed in the clinic, we used the term POC for both assays, in line with the following consensus definition: a 'point-of-care test…is a test to support clinical decision making, which is performed by a qualified…staff nearby the patient…during or very close to the time of consultation, to help the patient and physician to decide upon the best suited approach, and of which the results should be known at the time of the clinical decision making' (15).***'
Introduction	
Please revise to include information related to TV testing and the POC test for TV evaluated here. Again, only Xpert CT/NG is included.	As described above, we have added additional information on the OSOM TV assay to the Introduction and other sections. ***'Considering the relatively high cost of the Xpert Trichomonas vaginalis (TV) cartridge (~ USD 19.00), we decided to complement the Xpert CT/NG with the OSOM TV antigen detection assay (Sekisui, Lexington, MA, US) and Gram stain microscopy, in order to offer the participants a comprehensive 2-hour STI testing alternative to syndromic***

	management (10). The advantage of the OSOM TV assay is that it is relatively cheap (~ USD 8.00), has a rapid processing time (~ 10 minutes), and has shown higher accuracy than wet mount microscopy, especially in women (11, 12).'
Given that ***Treponema pallidum*** is mentioned in the opening paragraph, authors might consider a brief mention of how syphilis is detected in this setting. Are POC tests also used for syphilis at all here?	Although POC syphilis testing is practiced in some private healthcare facilities, it is not yet widely available in the government sector clinics. We have added a section on syphilis detection to the discussion section: '***We excluded pregnant women from the CAPRISA 083 cohort study, because they were referred for antenatal care, and were not an appropriate population to pilot the EPT intervention in. However, pregnant women may be an important population to offer POC testing to in the future (16, 17), and perhaps combine the testing model with a POC syphilis assay (18).***'
The authors note the potential existing infrastructure available as a result of TB testing with Xpert. How practically available would these same devices be for doing STI tests?	We have added the following sentence and reference to the Introduction: ***'Multi-disease testing for HIV viral load monitoring, early infant diagnosis and TB using the GeneXpert platform was found to be feasible in rural Zimbabwe (7).'***
Methods	
please include exclusion criteria here. Please add some more detail and references to support use of the comparison assays Anyplex II and the FTD STD9. Please state why these assays were chosen and considered as the gold-standard for these analyses? Please add some brief details on how Xpert CT/NG and OSOM were performed. Or provide reference to manufacturer instructions. These tests operate quite differently and readers may like to have some brief information on how they work (if word count allows).	There were no exclusion criteria other than mentioned (pregnancy and HIV positive status). We have added additional information to the Methods section on why these assays were chosen: ***'The Anyplex STI7 and FTD STD9 were chosen as confirmatory tests, because they are both CE marked and are commercially available in South Africa. For epidemiological purposes, these assays also provided the opportunity to determine the prevalence of sexually transmitted organisms not routinely screened for in surveillance studies.'*** To keep the manuscript brief, we have included the website links to the POC assays in the Methods section: All POC tests were processed **according to manufacturers' specification (www.sekisuidiagnostics.com/products/130-osom-trichomonas-test and www.cepheid.com/us/cepheid-solutions/clinical-ivd-tests/sexual-**

	health/xpert-ct-ng) by laboratory technologists with experience using the GeneXpert platform at the clinic laboratory, but no access to participant clinical data.
Data analysis - Please include information related to sample size calculation for this study.	We have added the following sentence to the Data analysis section: ***'The sample size was pre-determined to assess the primary outcome of the CAPRISA 083 cohort study, which assessed the reduction in genital tract pro-inflammatory cytokines after POC testing, immediate treatment and EPT among women diagnosed with STIs.'***
Patient and Public Involvement - The section needs further work, perhaps elaboration of the statements in first two sentences as currently adds little to the paper overall.	We have revised this paragraph, which is a requirement for *BMJ Open* submissions: **'*Patients and the public were not involved in the design of and recruitment to the study.*** However, the syndromic STI management approach in South Africa often leaves women untreated and return to clinics with recurrent symptoms and partner notification and treatment services are inadequate. These experiences by patients led to the study design, and the implementation of POC testing and expedited partner therapy in the clinic. Patients took part in focus group discussions and were able to provide feedback to the study team ***on their experiences with the POC STI testing model*** (9).**'**
Results	
Please add flow chart as per STARD guidelines. Is there any information to explain the 2 invalid results?	As suggested, we have added a study flow chart to the manuscript (see Figure 1). We have no further information to explain the two invalid results, but our proportion of invalid results is consistent with previous studies.
Is there any information on the **crossing point/thresholds** for the discordant CT or NG tests positive on the Xpert? Previous work referenced (Causer et al) suggested these discordants were perhaps related to low organism loads, at threshold of assay detection. It would be interesting to include this information from your study if available.	We have added the cycle thresholds to the manuscripts as suggested by Prof Peters above.
What was the time delay between specimen collection and references laboratory testing? How were specimens transported etc? Could this have impacted on specimen quality,	We have added some additional information about the timing of confirmatory testing: '…Eswab™ (Copan, Brescia, Italy) specimen, which was sent to the regional National Health Laboratory Services reference laboratory ***for***

contributing to the CT and NG discordants observed?	***DNA extraction*** and parallel testing on the Anyplex II STI-7 Detection assay (Seegene, Seoul, Korea) ***within 24 hours of sample collection according to Clinical and Laboratory Standards Institute (CLSI) requirements***.'
Discussion	
the first sentence includes evaluation of performance of both Xpert CT/NG and OSOM TV - however as noted earlier, this is not stated in the intro or abstract objectives. Please add accordingly.	We have expanded on the OSOM TV assay in the Abstract, Introduction and Discussion sections.
The CAPRISA protocol included with this submission mentions the Xpert CT/NG and TV assays, not the OSOM TV test evaluated here. Why was the Xpert TV assay not evaluated? This would seem like the most appropriate POC tests to use in this setting? Understand it perhaps was not available – however this should be noted in the discussion and/or intro as it is an obvious option as is now available.	As suggested, we have added this information to the Introduction (see above comments).
Final paragraph should address TV POC test also as this was included in the study reported here.	We have added additional information about the OSOM TV to the Discussion, and have modified the final paragraph: 'In conclusion, we found the Xpert CT/NG to be accurate when used at the point of care in a LMIC clinic, **and it was complemented well by the OSOM TV assay**.'

VERSION 2 – REVIEW

REVIEWER	Remco Peters University of Pretoria & Anova Health Institute, South Africa
REVIEW RETURNED	27-Nov-2018

GENERAL COMMENTS	Thank you - all comments have been addressed.

REVIEWER	Louise Causer Kirby Institute, UNSW Sydney, Australia
REVIEW RETURNED	28-Nov-2018

GENERAL COMMENTS	Thank you for the opportunity to review the revised manuscript. The authors have addressed all the issues raised by the reviewers. I only have 2 minor comments for consideration:

	1. The decisions to use the TV OSOM rather than the Xpert TV was noted to be driven by direct cost of the test, however the relative costs in-scaling up an STI screening program, including the costs and logisitics of a QA and training framework may make the Xpert TV assay a reasonable addition in future. 2. In the discussion (paragraph 3), regarding the high cycle thresholds, the authors note "This highlights that some caution is needed when interpreting high cycle threshold results, and that retesting may have to be considered." I would rather think that this points, not to the need to retest, but rather to highlight the limitations of comparing two or more highly accurate molecular based assays, that challenge the threshold limits of each.

VERSION 2 – AUTHOR RESPONSE

1. Consistent with Dr Causer's suggestion we have added the following sentences to the discussion section:

'It is important to note that the decision to use the OSOM TV rather than the Xpert TV was driven by the direct cost of the test. However, the relative costs of scaling up a STI screening program, including the costs and logistics of a quality assurance and training framework may make the Xpert TV assay still a reasonable addition in the future, even more so, if the company Cepheid decided to launch a CT/NG/TV multiplex cartridge.'

2. Consistent with Dr Causer's suggestion we have amended the specified section in the discussion:

'*This highlights that some caution is needed when interpreting high cycle threshold results****, and while*** *retesting may be considered****, it also highlights the limitations of comparing two or more highly accurate molecular based assays, that challenge the threshold limits of each other.'***

There were no additional comments.