PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	A mixed-methods evaluation of point-of-care hepatitis c virus RNA testing in a Scottish prison.
AUTHORS	Byrne, Christopher; Malaguti, Amy; Inglis, Sarah; Dillon, John

VERSION 1 – REVIEW

REVIEWER	Akiyama, Matthew Albert Einstein College of Medicine / Montefiore Medical Center, Medicine
REVIEW RETURNED	28-Oct-2022

GENERAL COMMENTS	This is study is an important contribution on HCV point of care viral load testing in Scottish prison. The study is well designed, well written, and the conclusions and limitations are accurately stated. Major comments - How were participants grouped in the pilot phase in terms of whether they received a PoC test or a conventional test? Was this done by facility, provider choice? This has important implications for the interpretation of the results because lead to imbalance in patient characteristics and therefore could impact more distal outcomes. To that end, did the authors compare characteristic of those tested with Xpert vs conventional testing (e.g. sex, substance use, OAT, etc.). If not, these would be important data to report on (could potentially be added to Table 1). - For Xpert failures were participants re-tested with conventional testing? Were attempts made to retest using Xpert? It does not appears this was reported. It would be helpful to understand if this was from an implementation standpoint and repeat testing would have implications on cost. Apologies if this was missed. - Were there any procedures in place for those who were released prior to or while on treatment. Such procedures (or absence of) should be stated in the background to better understand the context of those were lost to follow up. Minor comments - Please spell out HMP at first instance for those not familiar with the acronym - Figure 2 would be clearer if the groups were labeled with which type of test(s) they comprise (i.e PoC vs. conventional)

REVIEWER	Howell, Jessica Burnet Institute
REVIEW RETURNED	15-Dec-2022

GENERAL COMMENTS	Authors Byrne and colleagues have presented a neat mixed methods pilot study to explore the impact of introduction of PoC GenXpert RNA testing in a large maximum-security prison in Scotland, with a focus on the practicalities of implementation. This is highly topical as such programs are being implemented across prisons worldwide. The paper is nicely written and concise; the mixed methods approach is appropriate for an implementation science research question. The authors are to be commended on their study design and approach. I have some suggestions and questions for the authors to consider for their manuscript. General: Abstract: Results: -I would consider rather than stating that there were “70 determinants” identified in the qualitative study, make a broad statement about what key findings were (eg: characteristics of individuals were identified as major challenges, or something along those lines) -PoC was “costlier than conventional testing but sensitive to multiple factors”- I would perhaps reword “costlier” and leave out the reference to sensitivity for brevity - you don’t provide those data so this doesn’t add much to the results in the abstract. Introduction: - “It can cause long term negative health outcomes”- perhaps say what some of these are? - Line 5 page 6 and line 38 page 6, split sentences are a little confusing to read, could consider reordering Methods: - Cost analysis - This methodology is simple and this is highlighted in the limitations appropriately. There is mention of sensitivity to multiple factors a couple of times in the paper, but it is unclear how sensitivity analyses were conducted (and I could not find these data in the Results) - Qual: unusual to not use coding software- was there a reason? - Combining independent interviews with two people and a focus group with three participants means you are using two different methods with different inherent bias; this is compounded by the convenience sampling. These biases should be discussed in more detail. This really is a very small sample size and it is very unlikely that data saturation was reached given 70 factors were identified. - Triangulation in 20% of those sampled is really tiny in an already small sample- why not triangulate all 6 given the number was so small? This would improve confidence in the accuracy of your findings. - Statistical analysis: - There is insufficient information provided to know how the cox proportional hazards modelling was conducted and selection process for final parameters in the two models. Given the small sample size, a limited model comparing those treated and those not treated might be more suitable. I appreciate that the word count is tight, therefore you might like to place additional methodological data in a supplement.

	- “median survival” in the context it is written here is technically a correct term, but most would describe this as “time to treatment” Results: - For the reader, it might make more sense to present the pre and post pilot intervention data before presenting the overall data of numbers tested and treated overall, in order to answer your research question up front. Discussion: - A lot of additional qualitative results are presented in the discussion. You have a table in the Supp methods with greater detail on results of the qual work- I suggest expanding on your results in the Results section, then speaking more broadly in the Discussion section. - “low error rates”- it is worth noting that those error rates are not low, being higher than the Cepheid data would suggest. You could mention other data that have found similar error rates in the real world (there are some from Australia for example in the NSP setting) and how this is related in part to the learning curve of testers. Additionally, I was concerned about the 17% who had either an inadequate or inaccurate result recorded in their medical record- it would be worth expanding on this, particularly in light of some of the qual data presented (eg: result being on a bit of paper that then needs to be entered manually). This plus cepheid error rate meant it appeared that >30% of PoC tests were not valid due to test error or human recording error- I may have misinterpreted these data but would be worth clarifying. - I would not state that requiring too many approvals was a reason not to interview incarcerated clients. I would simply say the focus of the study was on healthworker implementation Conclusion - The conclusion again had the “sensitive to multiple factors” statement which is not based on presented data- perhaps could be omitted. Figure 2: - It was a little unclear at first read why there were three groups receiving RNA test outcomes- it would help the reader to label these as pre pilot, post pilot PoC and post pilot conventional testing (if I have interpreted this figure correctly)

VERSION 1 – AUTHOR RESPONSE

Reviewer 1

Dr. Matthew Akiyama, Albert Einstein College of Medicine / Montefiore Medical Center Comments to the Author:

This is study is an important contribution on HCV point of care viral load testing in Scottish prison. The study is well designed, well written, and the conclusions and limitations are accurately stated.

Thank you for recognising the contribution this work can make to the wider field.

Major comments

1. How were participants grouped in the pilot phase in terms of whether they received a PoC test or a conventional test? Was this done by facility, provider choice? This has important

implications for the interpretation of the results because lead to imbalance in patient characteristics and therefore could impact more distal outcomes. To that end, did the authors compare characteristic of those tested with Xpert vs conventional testing (e.g. sex, substance use, OAT, etc.). If not, these would be important data to report on (could potentially be added to Table 1).

Response: The patients were retrospectively grouped (i.e. artificially) for analysis purposes based on a) when they received their test (pre/post) and b) what test type they received (conventional/PoC). At the time of receiving their test, the choice of which test to use was based on usual practitioner/patient preferences. We have added a sentence to clarify this to Methods-Study Design (p.7). We did not test characteristic differences among patients across test types as there was little divergence in the cohort (all male, most treatment experienced, no cirrhosis, 9/10 PWID). There were differences in OAT receipt, which could theoretically impact on time to treatment, but you will note in Table 1 we only had this data for cases who received treatment. As the primary outcome/analysis includes people who did not initiate treatment (i.e. survived for longer, to censor), for whom we had very limited data, it was not possible to make comparisons adjusted for OAT receipt – this would require only including treated cases, which would bias the analysis.

2. For Xpert failures were participants re-tested with conventional testing? Were attempts made to retest using Xpert? It does not appear this was reported. It would be helpful to understand if this was from an implementation standpoint and repeat testing would have implications on cost. Apologies if this was missed.

Response: Participants whose Xpert result was not valid were re-tested, either using the GeneXpert or by conventional methods. The method depended on the nurse/patient preference at the time, so it varied. Of those who had invalid/error tests (n=26, among 20 patients), 15 patients had evidence of re-testing using the GeneXpert, consuming 18 Xpert assays (some repeat errors), while five patients had no evidence of re-testing with the GeneXpert (i.e. conventional blood draw was sent to local labs instead). So you are quite right that re-testing with Xpert assays impacted costs, to the tune of approximately £720. These data have been added to the manuscript pages 10 and (Results-primary outcomes-paragraph 1) and 13 (Results-Secondary outcomes-paragraph 1).

3. Were there any procedures in place for those who were released prior to or while on treatment. Such procedures (or absence of) should be stated in the background to better understand the context of those were lost to follow up.

Response: For clarity, no specific procedures over and above usual care were implemented alongside this test of change. However, Tayside has a well-developed network of community-based HCV testing/treatment pathways and, typically, people released from prison to the local community are appointed to attend a nurse-led outreach clinic if testing is required. (See this paper for more detail.) If treatment is ongoing, they are either a) discharged with the remainder of their medication, and appointed for SVR testing at the appropriate date, or b) the medication is transferred to their nominated community pharmacy, from which the remainder is dispensed on a schedule, and appropriate follow-up organised via the pharmacy. Those designated LTFU in the manuscript were either not engaged through these venues post-release, and/or moved out of Tayside to another area. A couple of lines on this have been added to the Background section on page 6 (paragraph 3).

Minor comments

1. Please spell out HMP at first instance for those not familiar with the acronym.

Response: Added to first instance in the Background.

2. Figure 2 would be clearer if the groups were labeled with which type of test(s) they comprise (i.e PoC vs. conventional)

Response: Agreed. We have revised this figure (and the figure legend) in line with this suggestion.

Reviewer 2

Dr. Jessica Howell, Burnet Institute

Comments to the Author:

Authors Byrne and colleagues have presented a neat mixed methods pilot study to explore the impact of introduction of PoC GenXpert RNA testing in a large maximum-security prison in Scotland, with a focus on the practicalities of implementation. This is highly topical as such programs are being implemented across prisons worldwide. The paper is nicely written and concise; the mixed methods approach is appropriate for an implementation science research question. The authors are to be commended on their study design and approach.

I have some suggestions and questions for the authors to consider for their manuscript.

Thank you for acknowledging the salience of the work and suitability of the study design.

Abstract

1. I would consider rather than stating that there were “70 determinants” identified in the qualitative study, make a broad statement about what key findings were (eg: characteristics of individuals were identified as major challenges, or something along those lines).

Response: This has been amended to read more generally and highlight the two most salient domains of the CFIR (page 3).

2. PoC was “costlier than conventional testing but sensitive to multiple factors”- I would perhaps reword “costlier” and leave out the reference to sensitivity for brevity - you don’t provide those data so this doesn’t add much to the results in the abstract.

Response: Agreed and amended (page 3).

Introduction

3. "It can cause long term negative health outcomes"- perhaps say what some of these are?

Response: We have added a few examples to the revised text on page 5 (paragraph 1).

4. Line 5 page 6 and line 38 page 6, split sentences are a little confusing to read, could consider reordering

Response: Amended as requested.

Methods

Cost analysis:

5. This methodology is simple and this is highlighted in the limitations appropriately. There is mention of sensitivity to multiple factors a couple of times in the paper, but it is unclear how sensitivity analyses were conducted (and I could not find these data in the Results)

Response: Apologies, this is a case of using terminology in the wrong context on our part. By sensitivity to multiple factors, we meant factors in the pathway which led to either 1) higher costs being incurred among those tested with the GeneXpert and 2) the higher LTFU rate among those tested with the GeneXpert. I.E., if fewer people tested with the GeneXpert started treatment, overall medication costs would have been lower and the cost-per-SVR would have been more favourable. Ditto for the LTFU rate, if this was lower in the GeneXpert group (or higher in the conventionally tested group), the cost-per-SVR would have been more favourable for the PoC patients. These were noted in the Discussion (page 20, paragraph 3).
We did not undertake a *sensitivity analysis* on the cost data which was subject to differing assumptions to the main analysis. The rudimentary methodology somewhat prohibits this (we did not use mathematical modelling as noted in the ms). For clarity, we have removed all mentions of *sensitivity* from the manuscript, where it refers to the costs, and used simpler phrasing. Thank you for highlighting.

Qual:

6. Unusual to not use coding software- was there a reason?

Response: Whilst coding software can be useful data management tools for larger collaborative qualitative research projects, particularly those which include multimedia, the need for software was discussed by team members and the decision was taken not to use one given the size and scope of this project. Also, in the view of the lead coder, the manual process facilitated engagement and familiarisation with, and interpretation of the data.

7. Combining independent interviews with two people and a focus group with three participants means you are using two different methods with different inherent bias; this is compounded by the convenience sampling. These biases should be discussed in more detail. This really is a very small sample size and it is very unlikely that data saturation was reached given 70 factors were identified.

Response: We understand your concerns and have expanded on the limitations of the approach taken in the Limitations section on pages 22-23. We noted in the manuscript on page 14 (Secondary outcomes, paragraph 2) that the individuals interviewed were almost all of the available sampling population. With regard to saturation, although often a target of qualitative research data collection, there are inherent uncertainties to it conceptually, particularly with the subjective processes employed to 'achieve' it (see Saunders and colleagues 2018; link). We interviewed a high proportion of the available sampling population (6/8; 75%) and consequently would suggest this might confer a reasonable degree of confidence in the results.

8. Triangulation in 20% of those sampled is really tiny in an already small sample- why not triangulate all 6 given the number was so small? This would improve confidence in the accuracy of your findings.

Response: the 20% triangulation strategy was agreed prior to the participants being approached and interviews undertaken, based on previous experience of the research team. The team agreed on a minimum of 80% concordance (in codes/CFIR allocation) to be achieved, indicating high inter-coder agreement. If not achieved, the pre-specified workflow meant triangulation would have been undertaken on additional transcripts. However, high concordance was achieved, and we therefore had confidence that remaining transcripts were accurately analysed. This strategy was the most efficient use of the resource available to the project, and meant initial triangulation at least *informed* the coding of all transcripts. We have added the strategy/workflow to the supplementary file (Figure S1, [Appendix S1].

Statistical analysis:

9. There is insufficient information provided to know how the cox proportional hazards modelling was conducted and selection process for final parameters in the two models. Given the small sample size, a limited model comparing those treated and those not treated might be more suitable. I appreciate that the word count is tight, therefore you might like to place additional methodological data in a supplement.

Response: The primary reasons for splitting the conventionally tested patients into two groups based on time (2018-19 & 2019-21) for inclusion in different models were two-fold: 1) to account for the effects of any changes that impacted BBV service delivery during the pilot period which were beyond our control (e.g. anything implemented by or impacting upon the prison service, e.g. staffing, changing reception processes); and 2) Covid-19, which occurred during the pilot and impacted upon turnaround time for conventional test results (generally speaking, faster pre- and slower post-) – this is evident in the median survival to treatment between the conventionally tested groups [p.11-12 beneath Table 1]). To improve confidence in any survival finding we wanted to ensure it would hold across both time periods, independently of impaired service delivery during the pilot. Realistically, model 1 probably gives a more realistic view of the differences in hazard as the conventionally tested group represents 'normal service delivery' which was unencumbered by Covid-19. We have added a line on this in the Discussion on page 19.

Both models were adjusted for age as there was a fairly wide range among cases included in the models (mean 38.9 SD 7.2; min 23 max 55) and, based on our experience, those in HMP Perth who are younger tend to be a little more transitory, which can impact upon the time horizon for healthcare engagement. That the impact of PoC testing held independently of age suggests the effect is not simply the by-product of nurses testing a less transient population. Age was the best available proxy for this as we did not have access to sentencing data.

However, in light of your suggestion, we have run an additional simplified model with patients grouped simply by test type (conventional vs PoC), without time period, both unadjusted and adjusted for age. These are included in the supplementary file (Table S1). You will see the PoC effect held in these additional analyses. We have amended the methods to reflect the decision-making process on page 8 (Methods-Statistics). We have also amended the Discussion on page 19 (paragraph one) to comment on this realism of the observed effect.

10. "median survival" in the context it is written here is technically a correct term, but most would describe this as "time to treatment"

Response: Happily amended in the revised version (page 12)

Results

11. For the reader, it might make more sense to present the pre and post pilot intervention data before presenting the overall data of numbers tested and treated overall, in order to answer your research question up front.

Response: If you are amenable, we would be grateful to be allowed to keep the current structure of the Results. We believe that giving the reader an awareness of the characteristics of the treated cohort (table 1) and an idea of how the number of treated cases in the survival analysis were arrived at (the preceding paragraphs) is helpful in contextualising the survival results. Dr Akiyama hinted at it in his comment 1: you can see, generally, that there are not many individual differences that might account for survival difference among patients beyond the test type (barring OAT status, data for which we did not have for untreated cases). Knowing this in advance of the primary outcome results, we feel, is useful.

Discussion

12. A lot of additional qualitative results are presented in the discussion. You have a table in the Supp methods with greater detail on results of the qual work- I suggest expanding on your results in the Results section, then speaking more broadly in the Discussion section.

Response: We had placed the table of determinants in a supplementary file for brevity in the main text. However, in light of this comment we have moved it to the main text (now Table 5). We would gently push back on the implication that new results are presented in the Discussion. There are no new data presented, no new quotations or tables, and the discussion points (and interpretation) speak generally about determinants now listed in Table 5 in the Results section, whilst referencing the CFIR domains that are listed in the left-most column of the table. Consequently, if amenable, we would like to retain this section of the Discussion as it is.

13. "low error rates"- it is worth noting that those error rates are not low, being higher than the Cepheid data would suggest. You could mention other data that have found similar error rates in the real world (there are some from Australia for example in the NSP setting) and how this is related in part to the learning curve of testers. Additionally, I was concerned about the 17% who had either an inadequate or inaccurate result recorded in their medical record- it would be worth expanding on this, particularly in light of some of the qual data presented (eg: result being on a bit of paper that then needs to be entered manually). This plus cepheid error rate meant it appeared that >30% of PoC tests were not valid due to test error or human recording error- I may have misinterpreted these data but would be worth clarifying.

Response: We have included a new table in the supplementary file (Table S2) which gives details of the result reporting errors. These would likely have been avoided if we had been able to implement an electronic result reporting system. We have expanded on this in the Discussion on page 22 (paragraph 6). We have also included a new figure in the supplementary file (Figure S2), which illustrates the number of failed tests over time. As you rightly noted, these generally decreased over time, which would imply an association with operator familiarity/proficiency. We have expanded on this in the Discussion on page 20 (paragraphs 1/2). We have cited two other studies (from Australia and Georgia) in the Discussion regarding error rates and revised the 'low' statement.

Finally, we have included another supplementary figure (Figure S3) which shows the number of result reporting errors over time, which spiked following remobilisation of services after the initial Covid outbreak. We added a comment on this in the Discussion on page 20 (paragraph 3).

14. I would not state that requiring too many approvals was a reason not to interview incarcerated clients. I would simply say the focus of the study was on health worker implementation

Response: Amended as requested (pages 22-23). Whilst you are right in saying the project had a health-systems implementation focus, it is worth saying that it was not simply a case of 'too many approvals'. The design of the project did impose certain limitations on scope of those who could be engaged.

Conclusion

15. The conclusion again had the "sensitive to multiple factors" statement which is not based on presented data- perhaps could be omitted.

Response: In line with previous response we have amended references to ‘sensitivity’ in the cost analysis, including here.

Figure 2

16. It was a little unclear at first read why there were three groups receiving RNA test outcomes- it would help the reader to label these as pre pilot, post pilot PoC and post pilot conventional testing (if I have interpreted this figure correctly)

Response: We have revised this figure in line with a similar suggestion from reviewer one, we hope this change also addresses this comment.

VERSION 2 – REVIEW

REVIEWER	Akiyama, Matthew Albert Einstein College of Medicine / Montefiore Medical Center, Medicine
REVIEW RETURNED	26-Jan-2023

GENERAL COMMENTS	The authors have done a great job in responding to reviewer comments. In light of their responsive revision, I feel the manuscript is ready for publication.

REVIEWER	Howell, Jessica Burnet Institute
REVIEW RETURNED	13-Feb-2023

GENERAL COMMENTS	The authors have addressed both my and Reviewer 1's questions and suggestions thoroughly. I feel the manuscript is now suitable for publication.