PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Patient outcomes in public sector Hepatitis C treatment programs: A retrospective cohort analysis across five low- and middle-income countries
AUTHORS	Boeke, Caroline; Adesigbin, Clement; Adisa, Olayinka; Agwuocha, Chukwuemeka; Akanmu, Muhammad-Mujtaba; Anartati, Atiek; Aung, Khin Sanda; Azania, Amy; Bello Nabe, Ruth; Budiman, Arief; Chan, Yuhui; Chawla, Umesh; Fatchanuraliyah, Fatchanuraliyah; Fernandes, Oriel; Grover, Gagandeep; Naing, Thandar Su; Ngo, Dang; Ramers, Christian; Regan, Sean; Sindhwani, Siddharth; Tandy, Gertrudis; Tint, Khin; Nguyen, Kinh; Witschi, Magdalena; McClure, Craig

VERSION 1 – REVIEW

REVIEWER	Miller, Lesley S. Grady Hlth Syst
REVIEW RETURNED	27-Apr-2022

GENERAL COMMENTS	General Comments This manuscript reports on a retrospective cohort analysis of hepatitis C treatment initiation and outcomes in four public sector HCV treatment programs in low- and middle-income countries, or LMIC (India, Myanmar, Nigeria and Vietnam). This study is of interest to BMJ Open readership and fills an important data gap as the study is a large, real-world study and reports HCV treatment outcomes from LMIC programs, on which there is limited data published. Overall, this paper was extremely well-written. The writing is clear and engaging and the paper is well-organized. Particular strengths of this paper include the authors' efforts to translate their findings into actionable, evidence-based strategies to improve the HCV treatment cascade. The authors also provide thoughtful insights into the study's limitations. This paper has the potential to contribute meaningfully to our understanding of the global HCV care cascade. Specific Comments Page 3, line 42: Please clarify the sentence starting with, "Given..." This is an important qualification regarding why all patients with HCV weren't included in the analysis and the sentence is confusing as written. Page 4, line 23: This is a helpful description of the standardized treatment algorithm and provides important context for the subsequent analysis.

	Page 4 (Table 1): In the “India” column, the year of program initiation is listed as 2018 but in the text later (page 5, line 56) the initiation year is listed as 2016. Also, in the row beneath (time period of data included), the start date is Feb 2016, which would mean data was collected before the start of the program. Also in Table 1, the last row defining treatment completion across programs is quite helpful. Page 5, line 45: Using a treatment start date cutoff of 6 months prior to data collection would allow time for SVR only for 12-week regimens, but not 24-week regimens. Page 6, paragraph starting on line 12: The authors clearly describe different strategies across programs for defining cirrhosis, but do not include how decompensated cirrhosis was defined. This would be a helpful addition. Page 6, line 32: It is quite helpful that the authors decided to analyze each country separately and adds to the validity of the results. Page 9, line 49: 5.4% overall seems like a very low prevalence with a history of IDU. It would be helpful to explore this further in the discussion. Page 10, line 20. It’s curious why such a high percentage of patients in Myanmar were prescribed 24 weeks of treatment, as it does not seem related to cirrhosis status. Page 15, line 35: This is an important limitation and the authors highlight it effectively Page 16, line 22: This is a key point, very well stated.

REVIEWER	Easterbrook, Philippa World Health Organization, Global Hepatitis Programme and HIV Department
REVIEW RETURNED	30-Apr-2022

GENERAL COMMENTS	• This is a well written paper with important findings regarding country programme experience with treatment completion, and cure. The authors have managed quite well the challenges of one country accounting for 95% of the data, and reporting the data separately by country. • • I have no major concerns. The main methodological issues include:- o It seems that the analysis of predictors on non-treatment completion, SVR tested, and cure rate were not adjusted. “Multivariate modeling was performed, but there were model convergence issues when accounting for data clustering by health facility, so unadjusted associations were reported in the final analyses.” o Gender and risk group are highly correlated in some countries ie most of PWID are male; and duration of regimen (24 weeks), type of regimen (with RBV) and presence of cirrhosis are also highly correlated. Some level of adjustment in these analyses should be undertaken and reported – so it is clear which are the core predictors. This could be done for Punjab.

	o Emphasise that Punjab India accounts for 95.6% of the study population. Surprisingly low number in Nigeria with 139 o Whenever India data is cited, best to state Punjab India, just as with Nasarawa State, Nigeria. Under objectives in abstract, important to emphasise that limited data on factors associated with treatment discontinuation and treatment failure in low and middle income countries. o Important to highlight that this data is from 3-4 years ago. Minor points • The term LTFU captures loss to follow-up prior to treatment, during treatment and after treatment. The term is more relevant to long-term monitoring, With hepatitis C, the issue is more treatment discontinuation. They may not be lost to follow, but they discontinued treatment. • Does not cite WHO data that 13% globally of those with chronic HCV infection initiated treatment. • Please clarify the difference between the 2020 paper of 7 programmes and this data from similar sites. The 2020 paper included Cambodia, Indonesia, Rwanda. There were four countries that were the same – India, Myanmar, Nigeria and Vietnam, and India accounted for most of them, and was based on 120,000. • Consider in Table 1: including number and % in the table. • But in Myanmar not completing treatment was defined by not returning for SVR12 – so not clear how it was possible to discriminate between “Completed treatment but did not return” 18.7%2 8.3% 58.3% 32.0% for SVR12 in Myanmar • Other questions: In terms of gender, was higher proportion of males in Vietnam due to high proportion that were PWID? Which patients were more likely to be cirrhotic – higher proportion in Myanmar and Vietnam? • • Regimens • “79.1% of patients were on sofosbuvir (SOF) and daclatasvir (DCV); all patients in Myanmar and Nigeria’s programs were prescribed SOF+DCV. In Vietnam, 37.3% of patients were prescribed SOF/DCV+ ribavirin (RBV) and in Punjab, India, 8.2% were prescribed sofosbuvir and velpatasvir (SOF/VEL) and 12.8% were prescribed SOF/VEL+RBV (these regimens were sometimes used to treat cirrhotic patients in these countries due to lower cost, whereas in Nigeria and Myanmar, cirrhotic patients were typically prescribed 24 weeks of SOF/DCV)” Misleading to state 78.8% had SOF/DCV regimen, as only in Punjab programme was there a SOF/VEL option. SOF/DAC plus RBV is not a different regimen, - it is still a SOF/DCV based regimen. Same applies to SOF/VEL and SOF/VEL +RBV

REVIEWER	Watson, Peter University of Cambridge, MRC Cognition and Brain Sciences Unit
REVIEW RETURNED	10-Jun-2022

GENERAL COMMENTS	Patient outcomes in public sector Hepatitis C treatment programs: A retrospective cohort analysis across four low- and middle-income countries bmjopen-2022-062745

	A few comments including on the describing of the places and people who administered the treatments, the reason for the missing data and to what outcomes they apply, how the vaguely described chi-square analysis compares to a more standard way of handling clustering and when it was able to be used and the influence of discrepancies in defining key outcomes such as positive diagnosis and completion of treatment. Page 3, line 22. Are the overall measures here averages of percentages averaged across countries e.g. you mention 79.1% were prescribed sofosbuvir+daclastasvir (line 23) but I am not sure how this percentage ties in with the percentages given on page 12, lines 52-56 in table 3 which appear to involve these drugs. Page 3, lines 25-29 Abstract. There are a series of p-values quoted here where it is not clear what two (or more?) proportions are being compared when a p-value is quoted. For example, on line 25 a p-value is given stating that males were less likely to complete treatment than females yet only a single proportion (88.4%) is quoted. You should quote both the completion proportions being compareed by the statistical test for males and females. Page 7, lines 12-25. There seems to be a range of different definitions about what constitutes a positive diagnosis (and on treatment completion) e.g. for cirrhosis APRI cut-offs are specified and in Myanmar were changed (in 2019). Fibroscans appear to be additionally used in some countries but not in others for diagnosis purposes. How confident, therefore, give these variations in diagnosis can we be in the authenticity of the diagnosis if there is so much disagreement about what constitutes a positive diagnosis? Page 7, lines 41-42. Could you elaborate here what you mean by the clusters which you describe as each representing a "health facility/treatment site" particularly the 59 treatment sites in Punjab state (page 6, line 55) since the clusters in the other countries are more fully described as being either hospitals or health centres (page 7 lines 3-10). Are these "treatment sites" places other than hospitals or health centres? Who administered the treatments? Were the people administering the treatment trained to do so? Page 7, line 42. Could you have used generalized linear mixed models here to adjust the comparison of proportions for clustering using e.g. the glmer procedure in R or genlinmixed in SPSS? This is the approach I have seen repeatedly used with clustered data and it can compare frequencies between clusters such as schools and classes where pupil scores are the outcome or repeated measures where subject is the cluster. I, therefore, wondered how the clchi2 procedure mentioned on line 42 differs, if at all, from a generalised linear mixed model approach with health facility as the cluster variable. Does this clchi2 procedure have a name or is to just known, as you mention here, rather vaguely as a chi-square adjusted for clusters? Is there a reference you can add in for this statistical method which looks at cluster adjusted percentage differences?

	Page 7, lines 42-43 and page 9, Table 2, page 12, Table 3, page 14, Table 4. I am not sure about putting patients with missing data in a separate category. Could you explain here in what analyses you then used this 'missing' category and how you interpret the people in this category. Are the people in the 'missing' category a diverse random set of individuals (missing completely at random) or are these people missing for various reasons some of which might have influenced the magnitude of their missing outcome (missing at random or even missing not at random)? Looking at Tables 2 (Page 9, line 35) and 3 (page 12, line 25) and also 4 (page 14, line 41) I only see missing data reported for age. Was there any other missing data on variables other than age? It is quite interesting you have missing ages since in studies, in the UK at least, we tend to obtain age and gender immediately at baseline by asking people what their ages are so don't tend to have any missing ages. I also wasn't sure what the dashes in the tables represented? Are these dashes referring to 0% or to unknown percentages? Page 7, lines 47-49. Could you explain which particular analyses precluded the use of adjustments for clustering? I assume the lack of convergence is due to small numbers of people in the various health facilities? Given the importance attached earlier the the paragraph to the adjustment for cluster effects are these analyses of percentage differences still appropriate if they have not been adjusted for cluster effects? Not sure what you mean by multivariate modeling (line 48). I assume you just compare each factor (e.g. sex) separately in univariate analyses on an outcome rate so the analysies would be univariate. Page 9, Table 2 and page 10 line 40, lines 49-50. Is it worth putting an extra column in Table 2 corresponding to overall percentages since you refer to an overall percentage (30.9% of patients being female) on page 10, line 40 and also to an overall percentage for HCV and drug use on page 10, lines 49-50. Page 11, line 44. I can in Table 3 see the 74.4% for decompensated cirrhosis in India reported in the text but don't see its comparator, 90.6% mentioned here for "across other categories" in India in Table 3. Similarly the 90.6% for across other ages in India is not in Table 3. Are these percentages obtained "across other categories" averages of the percentages across particular categories stated in table 3 and, if so, which categories? Page 14, line 9 and lines 29-30, Table 4. You mention cure rates in Vietnam (line 9) are lower in males than females yet on line 30 the percentage of cured males (99.5%) is greater than cured females (98.5% on line 29). You mention these differences (in cure rates between males and females and also in PWID) are lower despite not being statistically significant yet you still use statistical tests to assess these differences. I am not, therefore, sure you can say there are differences based on subjective perception if they are not (objectively) statistically significant unless the study is underpowered which then undermines the usefulness of statistical testing so I would remove references to differences in percentages if they do not meet statistical significance. I think as you say on page 16, line 35 we have to be careful in conclusions which may be influenced by confounding and one might quote specific examples of possible confounding e.g. by saying that any cure rates which differ due to sex may be, for example, due to females

	tending to be younger than males and if age is associated with cure rates it may be that the difference in cure rates with gender is due purely to age. Page 15, line 29, table 4. Given this footnote appears to refer to chi-square testing for individual age groups did you correct for multiple testing associated with testing the seven age groups separately with a combined group? Did you moreover test for a main effect of age ie do cure rates vary across the seven age groups in table 4 looked at together? I would be wary of testing individual age groups with the combined sample if there is no 'main effect' of age with cure rates? Given the very large sample sizes involved and the sensitivity of chi-square tests to large sample sizes any difference in percentages is likely to be statistically significant, however small they are. This should be mentioned as a caveat in the discussion. Where a statistical test is performed between two proportions one might consider more informatively quoting confidence intervals for the degree of difference being tested.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Dr. Lesley S. Miller, Grady Hlth Syst

Comments to the Author:
General Comments

This manuscript reports on a retrospective cohort analysis of hepatitis C treatment initiation and outcomes in four public sector HCV treatment programs in low- and middle-income countries, or LMIC (India, Myanmar, Nigeria and Vietnam). This study is of interest to BMJ Open readership and fills an important data gap as the study is a large, real-world study and reports HCV treatment outcomes from LMIC programs, on which there is limited data published. Overall, this paper was extremely well-written. The writing is clear and engaging and the paper is well-organized. Particular strengths of this paper include the authors' efforts to translate their findings into actionable, evidence-based strategies to improve the HCV treatment cascade. The authors also provide thoughtful insights into the study's limitations. This paper has the potential to contribute meaningfully to our understanding of the global HCV care cascade.
Response: We thank the reviewer for these comments.

Specific Comments

1. Page 3, line 42: Please clarify the sentence starting with, "Given..." This is an important qualification regarding why all patients with HCV weren't included in the analysis and the sentence is confusing as written.
Response: We have revised the language of this sentence to make it more clear: "Given that most countries only collected individual-level data for patients treated for HCV (with only aggregate-level data available on screening and diagnosis), this analysis focused on patients who initiated treatment."

2. Page 4, line 23: This is a helpful description of the standardized treatment algorithm and provides important context for the subsequent analysis.

Response: Thank you for this comment.

3. Page 4 (Table 1): In the "India" column, the year of program initiation is listed as 2018 but in the text later (page 5, line 56) the initiation year is listed as 2016. Also, in the row beneath (time period of data included), the start date is Feb 2016, which would mean data was collected before the start of the program. Also in Table 1, the last row defining treatment completion across programs is quite helpful.
Response: We initially set up Table 1 to cover the national-level programs, and India's national program started in 2018; however, we recognize that this may lead to confusion and given that our focus was Punjab, we have emphasized this at the top and changed the year to 2016, when Punjab's program started. We have added a footnote to clarify based on your question about the months in 2016: "the program was officially launched in June 2016, but a small number of patients initiated in the months prior."

4. Page 5, line 45: Using a treatment start date cutoff of 6 months prior to data collection would allow time for SVR only for 12-week regimens, but not 24-week regimens.
Response: This is a good point; we have revised the analysis to ensure that patients on 24-week regimens had adequate follow-up data to be included within each analysis. Of note, to maximize use of the data, for Table 2, we include *all* patients initiated on treatment, regardless of length of follow-up. We have revised the text to read: "Only patients with adequate follow-up time were included in analyses of patient treatment completion (at least 12 or 24 weeks to complete regimens as appropriate) and cure (at least 12 or 24 weeks to complete regimens as appropriate, plus an additional 12 weeks to complete SVR12 testing)."

5. Page 6, paragraph starting on line 12: The authors clearly describe different strategies across programs for defining cirrhosis, but do not include how decompensated cirrhosis was defined. This would be a helpful addition.
We have added this information by country to Table 1.

6. Page 6, line 32: It is quite helpful that the authors decided to analyze each country separately and adds to the validity of the results.
Response: Thank you for this comment. We have also been able to add data from Indonesia's national HCV treatment database, which we believe substantially strengthens this paper.

7. Page 9, line 49: 5.4% overall seems like a very low prevalence with a history of IDU. It would be helpful to explore this further in the discussion.
Response: This is a good point. While history of injecting drugs varies substantially across country contexts, we would have expected it to be higher in certain contexts including in Punjab. This may be due to underreporting by patients/underrecording by healthcare workers; we have added a sentence commenting on this in the limitations section of the discussion: "For instance, 7.0% of patients in the dataset reported history of injecting drugs, lower than might be expected; this could be due to underreporting by patients and/or inconsistent solicitation/recording of this information by healthcare workers."

8. Page 10, line 20. It's curious why such a high percentage of patients in Myanmar were prescribed 24 weeks of treatment, as it does not seem related to cirrhosis status.
Response: Yes, over time in Myanmar's program, it has been observed that a slightly higher proportion of patients were prescribed 24 weeks of treatment than were cirrhotic (in the time period included here, the difference was 32.7% vs. 25.8%). There was some degree of clinical discretion implemented by clinicians, and some debate on the APRI score cutoffs early in the program, which may explain the differences observed.

9. Page 15, line 35: This is an important limitation and the authors highlight it effectively
Response: Thank you for this comment.

10. Page 16, line 22: This is a key point, very well stated.
Response: Thank you for this comment.

Reviewer: 2
Dr. Philippa Easterbrook, World Health Organization

Comments to the Author:
- This is a well written paper with important findings regarding country programme experience with treatment completion, and cure. The authors have managed quite well the challenges of one country accounting for 95% of the data, and reporting the data separately by country.
-
- I have no major concerns. The main methodological issues include:-

1. It seems that the analysis of predictors on non-treatment completion, SVR tested, and cure rate were not adjusted. “Multivariate modeling was performed, but there were model convergence issues when accounting for data clustering by health facility, so unadjusted associations were reported in the final analyses.”

Gender and risk group are highly correlated in some countries ie most of PWID are male; and duration of regimen (24 weeks), type of regimen (with RBV) and presence of cirrhosis are also highly correlated. Some level of adjustment in these analyses should be undertaken and reported – so it is clear which are the core predictors. This could be done for Punjab.
Response: Thank you for your feedback. In response to this comment as well as feedback from Reviewer 3, we have updated the analytic methods to a multivariable-adjusted generalized linear model that accounts for clustering. All p-values are now adjusted for age, sex, and other factors that were statistically significant in at least one country. In the countries with smaller datasets, to get around model convergence issues, we dropped out variables that were too sparse to be included in the model. The detailed multivariable-adjusted odds ratios, 95% confidence intervals, and p-values can be found in Supplemental Tables 1 and 2, but to maintain a public health focus within the text, we still report percentages within each group and adjusted p-values.

2. Emphasise that Punjab India accounts for 95.6% of the study population. Surprisingly low number in Nigeria with 139
Response: We have been able to add data from 7,424 patients in Indonesia’s national HCV treatment program, which we believe substantially strengthens this paper. However, we agree that it is a limitation that the majority of the data still come from Punjab. We have updated the strengths and limitations box and the limitations paragraph in the discussion to note that 89% of the data came from Punjab, India.

3. Whenever India data is cited, best to state Punjab India, just as with Nasarawa State, Nigeria.
Response: We have added the state names more prominently throughout the manuscript to make this clear.

4. Under objectives in abstract, important to emphasise that limited data on factors associated with treatment discontinuation and treatment failure in low and middle income countries.
Response: We have added this information to the abstract.

5. Important to highlight that this data is from 3-4 years ago.
Response: The Indonesia data that we have added is more recent, as is the data from Punjab, but we agree that some of the data is older; we have added this point to the limitations section of the Discussion: "Programs were implemented differently across country contexts and some country data were limited to the very initial stages of the program, so it was difficult to make direct comparisons between countries; the older program data may be less relevant."

Minor points

6. The term LTFU captures loss to follow-up prior to treatment, during treatment and after treatment.
The term is more relevant to long-term monitoring, With hepatitis C, the issue is more treatment discontinuation. They may not be lost to follow, but they discontinued treatment.
Response: This is a good point; we have changed the terminology to treatment discontinuation throughout.

7. Does not cite WHO data that 13% globally of those with chronic HCV infection initiated treatment.
Response: Thank you; we have added this citation in the second paragraph of the Background.

8. Please clarify the difference between the 2020 paper of 7 programmes and this data from similar sites. The 2020 paper included Cambodia, Indonesia, Rwanda. There were four countries that were the same – India, Myanmar, Nigeria and Vietnam, and India accounted for most of them, and was based on 120,000.
Response: The 2020 paper described program initiation and scale-up using aggregate data only; it incorporated data across the cascade of care to show total volumes from screening to SVR12. The current paper includes more detailed analysis of patient-level data for countries with IRB approval, to assess patient characteristics in relation to treatment outcomes. We have clarified in our description of the previous paper that it included aggregate cascade of care data only.

9. Consider in Table 1: including number and % in the table.
Response: We have added the number of patients from each country within Table 1, as well as the percent coverage of public sector treatment data during the time period.

10. But in Myanmar not completing treatment was defined by not returning for SVR12 – so not clear how it was possible to discriminate between "Completed treatment but did not return" 18.7%2 8.3% 58.3% 32.0% for SVR12 in Myanmar
Response: In Myanmar, not completing treatment was defined by not coming for a visit post-treatment completion when an SVR12 test was ordered via a laboratory requisition form; there was a group of patients who did complete treatment and return for that follow-up visit, but did not then go on to complete an SVR12 test. We have removed the text "when an SVR12 test was ordered" from Table 1 to avoid confusion on this point.

11. Other questions: In terms of gender, was higher proportion of males in Vietnam due to high proportion that were PWID?
Response: Yes, 98.6% of those identified as PWID in Vietnam were male, so this likely contributed to the high proportion of males. However, even among those reported to be in the general population and not identified as PWID, 72.6% were male. We have added text to acknowledge this to the second paragraph of the Results: "This was reflective of a larger

proportion of males both getting tested for HCV and testing positive for chronic HCV and a higher proportion of PWIDs being male in those countries (>92%).”

12. Which patients were more likely to be cirrhotic – higher proportion in Myanmar and Vietnam?

Response: In Myanmar and Vietnam, cirrhotic patients were more likely in the first year of the program (likely “warehoused”, sicker patients), older, HBV positive, HIV negative, and not report a history of injecting drugs. In Myanmar, cirrhotic patients were more likely to be female, but in Vietnam they were more likely to be male. (Please note that these are observations based on the relative proportions of cirrhotic patients across groups but were not modeled to assess statistical significance.)

13. Regimens

“79.1% of patients were on sofosbuvir (SOF) and daclatasvir (DCV); all patients in Myanmar and Nigeria’s programs were prescribed SOF+DCV. In Vietnam, 37.3% of patients were prescribed SOF/DCV+ ribavirin (RBV) and in Punjab, India, 8.2% were prescribed sofosbuvir and velpatasvir (SOF/VEL) and 12.8% were prescribed SOF/VEL+RBV (these regimens were sometimes used to treat cirrhotic patients in these countries due to lower cost, whereas in Nigeria and Myanmar, cirrhotic patients were typically prescribed 24 weeks of SOF/DCV)” Misleading to state 78.8% had SOF/DCV regimen, as only in Punjab programme was there a SOF/VEL option. SOF/DAC plus RBV is not a different regimen, - it is still a SOF/DCV based regimen. Same applies to SOF/VEL and SOF/VEL +RBV

Response: We have updated the text to read: “In Myanmar, Nigeria, and Vietnam, all patients were prescribed sofosbuvir/daclatasvir (SOF/DCV) based regimens. In Indonesia, 79.6% were prescribed SOF/DCV based regimens and the remaining patients were on other regimens including sofosbuvir+simeprevir-based regimens and elbasvir/grazoprevir. In Punjab, India, 79.0% were prescribed SOF/DCV based regimens and 21.0% were prescribed sofosbuvir and velpatasvir (SOF/VEL) based regimens. In India and Vietnam, RBV was added to treatment for some cirrhotic patients due to lower cost, whereas in Nigeria and Myanmar, cirrhotic patients were typically precribed 24 weeks of SOF/DCV.”

Reviewer: 3
Dr. Peter Watson, University of Cambridge

Comments to the Author:
Patient outcomes in public sector Hepatitis C treatment programs: A retrospective cohort analysis across four low- and middle-income countries bmjopen-2022-062745

A few comments including on the describing of the places and people who administered the treatments, the reason for the missing data and to what outcomes they apply, how the vaguely described chi-square analysis compares to a more standard way of handling clustering and when it was able to be used and the influence of discrepancies in defining key outcomes such as positive diagnosis and completion of treatment.

1. Page 3, line 22. Are the overall measures here averages of percentages averaged across countries e.g. you mention 79.1% were prescribed sofosbuvir+daclastasvir (line 23) but I am not sure how this percentage ties in with the percentages given on page 12, lines 52-56 in table 3 which appear to involve these drugs.

Response: Thank you for your feedback. The overall measures are percentages across all patients in the dataset rather than averages across countries. Please note that we have been able to add data from 7,424 patients in Indonesia’s national HCV treatment program to strengthen this manuscript, so the summary numbers throughout the manuscript have been

updated. We have added to the abstract "Across all patients, …" to clarify this point. This matches the "All" column added to Table 2 per your comment #8 below.

2. Page 3, lines 25-29 Abstract. There are a series of p-values quoted here where it is not clear what two (or more?) proportions are being compared when a p-value is quoted. For example, on line 25 a p-value is given stating that males were less likely to complete treatment than females yet only a single proportion (88.4%) is quoted. You should quote both the completion proportions being compareed by the statistical test for males and females.
Response: We had originally only listed the proportions in the first group due to abstract word count constraints; with the updates made to the analysis and given the 5 separate country analyses, we now only have room within the abstract word limits to compare groups broadly (no specific proportions compared or p-values listed).

3. Page 7, lines 12-25. There seems to be a range of different definitions about what constitutes a positive diagnosis (and on treatment completion) e.g. for cirrhosis APRI cut-offs are specified and in Myanmar were changed (in 2019). Fibroscans appear to be additionally used in some countries but not in others for diagnosis purposes. How confident, therefore, give these variations in diagnosis can we be in the authenticity of the diagnosis if there is so much disagreement about what constitutes a positive diagnosis?
Response: As this is a retrospective analysis of public sector program, we were unable to verify patient cirrhosis status data; we agree that different countries have implemented different standards to designate cirrhosis status and in many cases, ultimately the clinician may use her or his discretion to make this determination. We have included a sentence in the Limitations section of the Discussions stating: "Data were retrospective, routinely collected program monitoring and evaluation data, so there may be issues with data accuracy and completeness affecting analyses."

4. Page 7, lines 41-42. Could you elaborate here what you mean by the clusters which you describe as each representing a "health facility/treatment site" particularly the 59 treatment sites in Punjab state (page 6, line 55) since the clusters in the other countries are more fully described as being either hospitals or health centres (page 7 lines 3-10). Are these "treatment sites" places other than hospitals or health centres? Who administered the treatments? Were the people administering the treatment trained to do so?
Response: To clarify, we have added the following information to this sentence about the site types in Punjab into Table 1: hospitals, HIV treatment centers, opioid substitution therapy centers, and prisons. We have specified the types of sites in the other countries as well.

5. Page 7, line 42. Could you have used generalized linear mixed models here to adjust the comparison of proportions for clustering using e.g. the glmer procedure in R or genlinmixed in SPSS? This is the approach I have seen repeatedly used with clustered data and it can compare frequencies between clusters such as schools and classes where pupil scores are the outcome or repeated measures where subject is the cluster. I, therefore, wondered how the clchi2 procedure mentioned on line 42 differs, if at all, from a generalised linear mixed model approach with health facility as the cluster variable.
Does this clchi2 procedure have a name or is to just known, as you mention here, rather vaguely as a chi-square adjusted for clusters?
Is there a reference you can add in for this statistical method which looks at cluster adjusted percentage differences?
Response:
We can certainly see the case for using regression modeling here, and in light of your thoughtful feedback as well as Reviewer 2's feedback that analyses should at a minimum be

adjusted for age and sex, we have shifted the analysis methodology to use generalized linear models that account for clustering. Although mixed models was a useful suggestion, we felt that the glm accounting for clustering was more in line with the original analysis agreed upon by our country counterparts. The glm models provide more statistical power compared to the clustered chi-square tests, so you will notice that additional variables are statistically significant in the updated analyses. This approach allowed us to present multivariable-adjusted associations and 95% confidence intervals. Given that we wouldlike to keep the public health focus of this manuscript and want to avoid very large tables, we have included the full model outputs as Supplemental Tables 1 and 2 and focused on the adjusted p-values output from those models in the reporting in the text.
With regards to your question on the cluster-adjusted chi-square: More information on this statistical method can be found in Jung S., Ahn C., and Donner A. Evaluation of an adjusted chi-square statistic as applied to observational studies involving clustered binary data. Statist Med 2001; 20(14): 2149-61. doi: 10.1002/sim.857
We have frequently used this procedure for cluster-adjusted comparison of proportions across groups, and originally chose to use it in this analysis for ease of interpretation given that our focus here was public health messaging. This nonparametric procedure does not make assumptions about the data and therefore has more limited statistical power to detect differences between groups compared to parametric methods.

6. Page 7, lines 42-43 and page 9, Table 2, page 12, Table 3, page 14, Table 4. I am not sure about putting patients with missing data in a separate category. Could you explain here in what analyses you then used this 'missing' category and how you interpret the people in this category. Are the people in the 'missing' category a diverse random set of individuals (missing completely at random) or are these people missing for various reasons some of which might have influenced the magnitude of their missing outcome (missing at random or even missing not at random)?
Looking at Tables 2 (Page 9, line 35) and 3 (page 12, line 25) and also 4 (page 14, line 41) I only see missing data reported for age. Was there any other missing data on variables other than age? It is quite interesting you have missing ages since in studies, in the UK at least, we tend to obtain age and gender immediately at baseline by asking people what their ages are so don't tend to have any missing ages. I also wasn't sure what the dashes in the tables represented? Are these dashes referring to 0% or to unknown percentages?
Response: While there are many approaches to handling missing data (e.g., imputation, exclusion, etc.), we chose to analyze patients with missing data using the "missing indicator" approach. The missing indicator approach is a commonly used method to account for missing data in a multivariate regression model whereby for each variable in the model, a "missing" dummy variable (0/1) is created that includes the patients that are missing data for that variable. See Blake H.A., et al. Estimating treatment effects with partially observed covariates using outcome regression with missing indicators. Biochem J 2020; 62(2):428-443. DOI: 10.1002/bimj.201900041
Each method to account for missing data has strengths and limitations; while there is some risk of bias with the missing indicator method depending on the randomness of the missing data, this method was chosen because it is a simple and intuitive approach that uses all available data, thereby reducing loss of statistical power. We have added the following text to clarify the methods used and rationale: "A missing indicator category was included for each variable in the model to account for missing data and minimize loss of statistical power due to missingness. Results were similar when those with missing data were removed from analyses."
It is not possible to confirm whether those missing age were missing at random, but as shown in Table 2, age was only missing for 0.2% of patients in India, 2.2% of patients in Indonesia, and 0.1% of patients in Myanmar (and 0% in Nigeria and Vietnam), so missing age should not substantially impact conclusions drawn in the analysis. In Tables 3 and 4, we show

the percentage of patients within each category who completed treatment and who were cured, respectively; the dashes represent categories with no patients (i.e., the denominator of those cells is 0, hence a percentage was not calculated).
Missing/unavailable data for other variables is listed in Table 2.

7. Page 7, lines 47-49. Could you explain which particular analyses precluded the use of adjustments for clustering? I assume the lack of convergence is due to small numbers of people in the various health facilities? Given the importance attached earlier the the paragraph to the adjustment for cluster effects are these analyses of percentage differences still appropriate if they have not been adjusted for cluster effects? Not sure what you mean by multivariate modeling (line 48). I assume you just compare each factor (e.g. sex) separately in univariate analyses on an outcome rate so the analyses would be univariate.
Response: All p-values presented accounted for clustering, and with the adjustments to the analyses, all p-values presented are also multivariable-adjusted based on GLM regression models. We have revised the Methods section to reflect these changes.

8. Page 9, Table 2 and page 10 line 40, lines 49-50. Is it worth putting an extra column in Table 2 corresponding to overall percentages since you refer to an overall percentage (30.9% of patients being female) on page 10, line 40 and also to an overall percentage for HCV and drug use on page 10, lines 49-50.
Response: This is a good suggestion to ensure clarity. We have added this column to Table 2.

9. Page 11, line 44. I can in Table 3 see the 74.4% for decompensated cirrhosis in India reported in the text but don't see its comparator, 90.6% mentioned here for "across other categories" in India in Table 3. Similarly the 90.6% for across other ages in India is not in Table 3. Are these percentages obtained "across other categories" averages of the percentages across particular categories stated in table 3 and, if so, which categories?
Response: With the changes to the analysis methodology, in the updated manuscript we have changed the way that results are presented to compare specific categories within each variable to a reference category (rather than one category against all others, as we had done previously).

10. Page 14, line 9 and lines 29-30, Table 4. You mention cure rates in Vietnam (line 9) are lower in males than females yet on line 30 the percentage of cured males (99.5%) is greater than cured females (98.5% on line 29). You mention these differences (in cure rates between males and females and also in PWID) are lower despite not being statistically significant yet you still use statistical tests to assess these differences. I am not, therefore, sure you can say there are differences based on subjective perception if they are not (objectively) statistically significant unless the study is underpowered which then undermines the usefulness of statistical testing so I would remove references to differences in percentages if they do not meet statistical significance. I think as you say on page 16, line 35 we have to be careful in conclusions which may be influenced by confounding and one might quote specific examples of possible confounding e.g. by saying that any cure rates which differ due to sex may be, for example, due to females tending to be younger than males and if age is associated with cure rates it may be that the difference in cure rates with gender is due purely to age.
Response: Thank you for catching this typo on p. 14, line 9; we have fixed this text and now note "the opposite was true in Vietnam." We have removed observations of qualitative differences between percentages where not the p-values were not statistically significant, and all p-values presented are now multivariable-adjusted.

11. Page 15, line 29, table 4. Given this footnote appears to refer to chi-square testing for individual age groups did you correct for multiple testing associated with testing the seven age groups

separately with a combined group? Did you moreover test for a main effect of age ie do cure rates vary across the seven age groups in table 4 looked at together? I would be wary of testing individual age groups with the combined sample if there is no 'main effect' of age with cure rates?
Response: The updated analysis uses multivariable-adjusted regression to assess difference between specific categories of each variable. Given that this is real-world observational data with measurement error, that the independent variables are to some extent correlated with each other, and that potential associations are hypothesis-driven (i.e., not truly independent in the way that genetic wide association studies are), we did not correct for multiple testing in this analysis. Corrections for multiple testing such as the Bonferroni correction reduce the chances of Type I error at the expense of Type II error, and given the factors described above, we did not feel the use was justified for this analysis.
See Armstrong R.A. When to use the Bonferroni correction. Ophthalmic Physiol Opt 2014; 34(5): 502-8. doi: 10.1111/opo.12131
We agree that given the large sample size, particularly for Punjab, it is possible that some comparisons may have been statistically significant due to chance, but given that our goal was to identify groups that *may* benefit from additional support to complete treatment and achieve cure, we felt that it was justifiable to report all statistically significant findings as the more client-centered approach. We have added a sentence to the Limitations section noting this: "Given the large sample sizes and high level of statistical power, it is possible that some findings were statistically significant due to chance; we focused on risk factors identified as significant across multiple countries or outcomes and those with larger relative differences in outcomes, indicating greater public health relevance."

12. Given the very large sample sizes involved and the sensitivity of chi-square tests to large sample sizes any difference in percentages is likely to be statistically significant, however small they are. This should be mentioned as a caveat in the discussion. Where a statistical test is performed between two proportions one might consider more informatively quoting confidence intervals for the degree of difference being tested.
Response: With the updated analysis methods, we have included 95% confidence intervals in Supplemental Tables 1 and 2. As noted in our response to your Comment #11, we agree that given the large sample size, particularly for Punjab, some comparisons may have been statistically significant due to chance, but given that our goal was to identify groups that *may* benefit from additional support to complete treatment and achieve cure, we felt that it was justifiable to report all statistically significant findings. We have added a sentence to the Limitations section noting this: "Given the large sample sizes, some statistically significant findings may have been due to chance."

VERSION 2 – REVIEW

REVIEWER	Miller, Lesley S. Grady Hlth Syst
REVIEW RETURNED	28-Sep-2022

GENERAL COMMENTS	The authors have thoughtfully and thoroughly addressed the reviewers' comments, and the manuscript is strengthened as a result. My only additional suggestion is related to the "lost to follow up" vs. "treatment discontinued" terminology that was updated in this version. I'm not sure these can be used interchangeably, as a person may not return for treatment visits or SVR testing (ie lost to follow up) but could still have completed their medication course (ie not treatment discontinued). It may be helpful to define early on

	what outcomes are included under treatment discontinuation, ie define as "patients who either d

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1
Dr. Lesley S. Miller, Grady Hlth Syst
Comments to the Author:
The authors have thoughtfully and thoroughly addressed the reviewers' comments, and the manuscript is strengthened as a result. My only additional suggestion is related to the "lost to follow up" vs. "treatment discontinued" terminology that was updated in this version. I'm not sure these can be used interchangeably, as a person may not return for treatment visits or SVR testing (ie lost to follow up) but could still have completed their medication course (ie not treatment discontinued). It may be helpful to define early on what outcomes are included under treatment discontinuation, ie define as "patients who either discontinued medication or were lost to follow up" or something similar.

Response:

Thank you for this feedback. In response, in the background section (last paragraph, first sentence), we have added “treatment discontinuation (defined differently across country contexts as not collecting all medication refills or being lost to follow-up)”. In addition, we have highlighted in the data analysis section that country-specific definitions of treatment completion can be found in Table 1.

We hope that you will now find the manuscript suitable for publication.

Sincerely,
Caroline Boeke, ScD
Senior Technical Advisor
Clinton Health Access Initiative (CHAI)