PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Predictive Value of Clinician Impression for Readmission and Post-Discharge Mortality among Neonates and Young Children in Dar es Salaam, Tanzania and Monrovia, Liberia
AUTHORS	Rees, Chris Kisenge, Rodrick Ideh, Readon Kamara, Julia Coleman-Nekar, Ye-Jeung Samma, Abraham Godfrey, Evance Manji, Hussein Sudfeld, Christopher Westbrook, Adrianna Niescierenko, Michelle Morris, Claudia Whitney, Cynthia Breiman, Robert Manji, Karim Duggan, Christopher

VERSION 1 - REVIEW

REVIEWER	Dr. Kirby Tickell University of Washington
REVIEW RETURNED	12-Apr-2023

GENERAL COMMENTS	This is a simple but very informative study that will be helpful in the ongoing discourse about how to identify children at risk of poor outcomes in the post-discharge period. The question being addressed is clearly outline, the methods are appropriate to answering the question, those methods have been accurately implements, and the interpretation of the results seems to be very reasonable. The only very small note I have is that on Page 7, the sentence beginning on line 47 says "the most common discharge diagnoses" twice. Otherwise I cannot see any revision that need to made prior to publication.

REVIEWER	Dr. Peter Flom Peter Flom Consulting
REVIEW RETURNED	25-Apr-2023

GENERAL COMMENTS	I confine my remarks to statistical and methodological aspects of this article. Unfortunately there is a fairly severe problem: Using AUC and ROC when the data are skewed, as they are here, is not a good

	method. See https://towardsdatascience.com/imbalanced-data-stop-using-roc-auc-and-use-auprc-instead-46af4910a494 Intuitively, when very few cases are in one class (here, very few babies are readmitted) the curve is necessarily odd. In this paper, the fact that the clinician probabilities were strongly associated with outcome (see table 2 and others) is a further indication that something is wrong with the AUC analysis. Minor points: p. 6 In addition to giving the number of child deaths is subSaharan Africa, and the fact that this is half the worldwide total, the authors should say what % of children live in the region. I'm sure it's not 50% (that is, the risk in this region is elevated) but, without the %, we can't tell how much it is elevated. line 15 ff Maybe it is obvious to experts, but I think the authors should spell out why the tools used in high income countries wouldn't work in SS Africa. p. 8 If the authors do more work like this, I would recommend that the responses be either 0 to 100, or similar, rather than categorized. Also, why have 100% as a separate category? (From results, it seems like no one checked 100%, so, this didn't wind up making any difference). Table 2 Rather than have columns for total and readmitted, it would be better to have NOT readmitted and admitted. (You can have total, as well, if you want). This would make it much easier for readers to compare the relevant percentages. Also, the %ages listed make no sense. The 'total'' should always be 100%. This makes me think that the problem may be mislabeling of the columns. Something is wrong here, though. Table 4 Same issues as table 2 Overall recommendation: The analysis I would do here is logistic regression, with either readmission or mortality as the dependent variable, and clinician rating as one independent variable, but also have IVs for type of clinician (this is clearly important), age of child (in months) and other relevant variables.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1

This is a simple but very informative study that will be helpful in the ongoing discourse about how to identify children at risk of poor outcomes in the post-discharge period. The question being addressed is clearly outline, the methods are appropriate to answering the question, those methods have been accurately implements, and the interpretation of the results seems to be very reasonable.

Author response: We appreciate the reviewer’s comments and agree that our findings move this area of investigation forward.

The only very small note I have is that on Page 7, the sentence beginning on line 47 says "the most common discharge diagnoses" twice. Otherwise I cannot see any revision that need to made prior to publication.

Author response: We have adjusted the language in the second paragraph of the Results as suggested. This now reads, “The most common discharge diagnoses among neonates were sepsis (29.7%, n=609), prematurity (28.8%, n=591), and birth asphyxia (15.8%, n=323). Among infants and children, pneumonia (12.1%, n=223), diarrheal disease (10.1%, n=186), and malaria (7.2%, n=133) were most common.”

Reviewer 2

I confine my remarks to statistical and methodological aspects of this article. Unfortunately there is a fairly severe problem: Using AUC and ROC when the data are skewed, as they are here, is not a good method. See https://nam11.safelinks.protection.outlook.com/?url=https%3A%2F%2Ftowardsdatascience.com%2Fimbalanced-data-stop-using-roc-auc-and-use-auprc-instead-46af4910a494&data=05%7C01%7Cchris.rees%40emory.edu%7C6feb7266d8b34ad1e1e808db46fb3e1a%7Ce004fb9cb0a4424fbcd0322606d5df38%7C0%7C0%7C638181817178555507%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C3000%7C%7C%7C&sdata=DdytiMZwqIsf6av7JMh0ZNFZ%2B%2B%2BPPBnmG14GtLIKPWo%3D&reserved=0

Intuitively, when very few cases are in one class (here, very few babies are readmitted) the curve is necessarily odd.

Author response: We appreciate the reviewer’s comments and willingness to share insightful sources. After reading this article, we read further on precision-recall curves and believe this is an appropriate metric to use for our dataset. We have removed the AUC information and have instead estimated the area under the precision-recall curve (AUPRC) and 95% confidence intervals using 5-fold cross-validation according to Saito and Rehmsmeier (reference below). In almost all cases, the AUPRC closely approximated the prevalence of the outcome in that group indicating that clinicians’ identification of at-risk children performed similar to random chance within that group. When evaluating readmission, medical officers did perform better than chance in all age groups (AUPRC=0.23 [95% CI: 0.17-0.34] vs Chance=0.09) and in children only (AUPRC=0.30 [95% CI: 0.18-0.41] vs Chance=0.10). Overall, our conclusions did not change with these updated analyses although we believe the new analyses are stronger and an improvement to the prior version.

Saito T, Rehmsmeier M (2015) The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. PLOS ONE 10(3): e0118432. https://doi.org/10.1371/journal.pone.0118432

In this paper, the fact that the clinician probabilities were strongly associated with outcome (see table 2 and others) is a further indication that something is wrong with the AUC analysis.

Author response: After adding the non-readmitted and non-mortality columns to Tables 2 and 4, we believe it is clearer as to why there was an association between clinician probability and the outcomes that was not apparent in the AUC. The association between readmission probabilities and readmission is opposite of what we would expect in that the non-readmitted participants were typically given higher probabilities than those that were readmitted. However, participants who experienced mortality were more likely to be given a higher probability than those who did not experience mortality although there was still poor precision in detecting participants at risk for mortality. In either case, we

have removed the AUC analysis and replaced it with AUPRC which is more appropriate for an imbalanced dataset as you have kindly pointed out.

Minor points:

p. 6 In addition to giving the number of child deaths is subSaharan Africa, and the fact that this is half the worldwide total, the authors should say what % of children live in the region. I'm sure it's not 50% (that is, the risk in this region is elevated) but, without the %, we can't tell how much it is elevated.

Author response: We appreciate the reviewer's comment. In response, we have changed the reporting in the Introduction to reflect the disparate rates in mortality among children aged <5 years. The first sentence now reads, "Mortality rates among children aged <5 years in sub-Saharan Africa are the highest in the world at 74 per 1,000 live births, which is 14 times higher than rates in Europe and North America."

line 15 ff Maybe it is obvious to experts, but I think the authors should spell out why the tools used in high income countries wouldn't work in SS Africa.

Author response: We have added the following sentence to paragraph two of the Introduction in response to the reviewer's comment.

"Differences in healthcare access and disease prevalence in the United States and sub-Saharan Africa may necessitate the creation of clinical prediction rules catered to settings in sub-Saharan Africa."

p. 8 If the authors do more work like this, I would recommend that the responses be either 0 to 100, or similar, rather than categorized. Also, why have 100% as a separate category? (From results, it seems like no one checked 100%, so, this didn't wind up making any difference).

Author response: We appreciate the reviewer's comment and will certainly adapt accordingly in future work.

Table 2 Rather than have columns for total and readmitted, it would be better to have NOT readmitted and admitted. (You can have total, as well, if you want). This would make it much easier for readers to compare the relevant percentages.

Author response: We appreciate your thoroughness and have since added the non-readmitted and non-mortality columns to Tables 2 and 4, respectively.

Also, the %ages listed make no sense. The 'total'' should always be 100%. This makes me think that the problem may be mislabeling of the columns. Something is wrong here, though. Table 4 Same issues as table 2

Author response: We appreciate the opportunity to correct the typo in these tables. These have been fixed and triple-verified for both Tables 2 and 4.

Overall recommendation: The analysis I would do here is logistic regression, with either readmission or mortality as the dependent variable, and clinician rating as one independent variable, but also have IVs for type of clinician (this is clearly important), age of child (in months) and other relevant variables.

Author response: Thank you for the recommendation. As stated in the response above to the Associate Editor, we have incorporated binary logistic regression models with readmission and mortality as the outcomes into the article. Due to small sample sizes in the non-0% clinician probability categories, we reduced the categorization of clinician probability to 0%, 1-20%, 21-60%, and 61-99% for the models. We were able to adjust for physician type, whether the diagnosis was of an infectious nature or not, age at discharge, and duration of hospital stay. We attempted to adjust for

the reason of clinician probability with both the original categories (No risk, Clinician perceived inability to pay for treatment, Clinician perceived social concerns, Clinician perceived progression of illness, and Other) and as a dichotomous variable (Clinician perceived inability to pay for treatment versus No clinician perceived inability to pay for treatment) but due to collinearity with the perceived level of risk as identified with variance inflation factor, we were unable to include this predictor in the model. The following sentences have been added to the statistical analyses methods to reflect this newly added analysis:

“Additionally, we conducted binary logistic regression analyses to assess whether the perceived level of risk for each outcome was associated with the patient’s likelihood of each outcome after adjusting for clinician type, patient age at discharge (months), whether the discharge diagnosis was infectious or not, and duration of hospitalization (days). Due to small sample sizes in the non-0% clinician probability categories, we reduced the categorization of clinician probability to 0%, 1-20%, 21-60%, and 61-99%. The clinician cited reason for the outcome was also considered in the model but was removed due to collinearity as assessed by the variance inflation factor (VIF).”

The results from these models (as shown in Table 7) indicate that level of clinician probability, after adjusting for potential confounders, was not associated with mortality. However, those who were identified as having a risk between 21-60% had 0.45 times (95% CI: 0.26-0.74, p-value=0.003) the odds of being readmitted compared to those with a 0% clinician probability. This is in contrast with what we would expect to find, which would be that a higher clinician probability is associated with higher odds of the outcome occurring. The results from this model in conjunction with the results of the AUPRC analyses indicate that clinicians were not accurately able to identify at risk neonates and children.

VERSION 2 – REVIEW

REVIEWER	Dr. Peter Flom Peter Flom Consulting
REVIEW RETURNED	27-May-2023

GENERAL COMMENTS	The authors have addressed my concerns and I now recommend publication.

REVIEWER	Dr. Kirby Tickell University of Washington
REVIEW RETURNED	01-Jun-2023

GENERAL COMMENTS	Thank you for making these revisions.

VERSION 2 – AUTHOR RESPONSE

N/A