nature portfolio

Peer Review File

Respiratory syncytial virus hospitalisation by chronological month of age and by birth month in infants

Corresponding Author: Professor You Li

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Attachments originally included by the reviewers as part of their assessment can be found at the end of this file.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
This is a very well written manuscripts with interesting data presented. The prediction tool in this study could be very useful to countries without age-granular RSV disease burden data, particularly LICs and LMICs, to inform decisions on timing and prioritization of RSV immunization programs.

Here are some comments for the authors to consider:
1. Can the authors consider using formulae notations (these can be included as supplemental materials) for the various calculations? This should make it easy to understand and follow the methods used, including reproducing similar analyses for those who might be interested in doing so.

2. Consider adding a box in Fig 2 to show the study that was excluded from the quantitative analysis because of report of pooled results.

3. Could the authors provide a rationale, either in the introduction or discussion, for why they also analyzed data by latitudinal group (temperate, sub-tropical and tropical)?

4. For the online tool described in lines 258 - 259, I think it would be useful for the authors to allow the user to optionally use percent RSV positivity data as “seasonality data” if the country thinks that case counts could be biased or misleading. Using only cases, as provided for in the prediction tool, requires that testing of patients for RSV is done systematically all through the year. For countries such LICs and LMICs, who could potentially benefit most from this tool, testing for RSV might be conducted differentially for various reasons, including limited testing supplies.

5. In line 293, did the authors mean to say that “a country could include multiple datasets”? It currently reads “study”, yet my understanding is that each dataset is a study.

6. Might the authors consider showing the data from the ad-hoc verification analysis (described in lines 315 - 320) in a graphical presentation as a supplemental figure similar to Fig 3? This could just be one graph each (for the 134 vs 54 studies) and not necessarily by income status.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
My comments to the authors are in the attached pdf.

(Remarks on code availability)

As far as I can tell, the authors have not made any of their code available.

Reviewer #3

(Remarks to the Author)
This is an elegant study determining the risk of respiratory syncytial virus (RSV) hospitalization using two important and readily available risk factors: chronological age and birth month. The authors developed a country/latitude specific prediction tool predicting country specific RSV hospitalization burden. The only information that needs to feed into the prediction tool is monthly RSV surveillance data. The prediction tool can be of significant public health importance, guiding country/region specific RSV immunization programs and improving vaccine efficiency. My main comments are:
1. Validity of the underlying assumption of independence of chronological age and birth month on the risk of RSV hospitalization. The prediction tool is developed with this underlying assumption. Throughout the manuscript, the relationship between these two factors and the risk of infant RSV hospitalization is presented separately, though at the same time, interpreted as there is an interaction effect. It is confusing whether the effect of the two factors is truly independent or not. The varying cumulative risk in infancy due to birth month hints an interaction effect (supplemental figure).
2. The RSV hospitalization burden presented in the manuscript is relative to the overall infancy RSV hospitalization (proportions). Infant RSV hospitalization prevalence itself varies by countries. It might be even more policy impactful to "predict" country specific month-by-month absolute number of RSV hospitalizations and the associated costs for each birth month.
3. The study focuses on RSV hospitalization, it is important to note that RSV hospitalization is only a small fraction of infant medically attended RSV illnesses. Infant RSV associated disease burden will be higher
4. Limiting the risk factor to chronological age and birth month is the strength of the study, it also limits the application of the study/prediction tool to more public health policy level as many subject specific risk factors such as preterm birth, Down syndrome, chronic lung disease due to prematurity… are not included in the prediction tool.

(Remarks on code availability)
The web link of the prediction tool is very user friendly and self explanatory

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
Thanks for your response. My comments have sufficiently been addressed.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
1) Response to my first comment: in my opinion, the model the authors are using to model the number of hospitalizations is not appropriate. According to their response, they are assuming that all observations within the same income stratum are independent. This is not realistic, as it is much more likely that observations from the same study are more similar to one another than observations from different studies. Put another way, observations from the same study are correlated. While this issue of correlation will likely not have an impact on point estimates (e.g., their solid lines in Figure 2), the credible intervals they are reporting are likely much too narrow.

The authors can account for within-study dependence using the first model I proposed in Comment 1 of my original response. They should also be accounting for birth month if they would like credible intervals to be believable (see below).

2) Response to my second comment: unfortunately, the authors did not address my concern. I had concerns about the within-study dependence caused by not accounting for birth month in their proposed model, not their stated objectives (which are excellent!) As noted in my second comment, this dependence implies their credible intervals (Figures 2 and 3) are too narrow. If the authors cannot account for birth month in their model (whose estimates are used to generate Figures 2 and 3) or cannot show that birth month has a negligible impact on hospitalizations, I would recommend removing credible intervals entirely or state that they are likely too narrow.

(Remarks on code availability)
The code looks great.

Reviewer #3

(Remarks to the Author)
All my comments are well addressed.

(Remarks on code availability)

GENERAL RESPONSE

We thank the reviewers and the editorial team for their time on evaluating our manuscript. We have carefully considered the comments and revised our manuscript accordingly. Herewith we provide a point-by-point response to each of the comments.

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

This is a very well written manuscripts with interesting data presented. The prediction tool in this study could be very useful to countries without age-granular RSV disease burden data, particularly LICs and LMICs, to inform decisions on timing and prioritization of RSV immunization programs.

Here are some comments for the authors to consider:

Comment #1

1. Can the authors consider using formulae notations (these can be included as supplemental materials) for the various calculations? This should make it easy to understand and follow the methods used, including reproducing similar analyses for those who might be interested in doing so.

Response: We appreciate the suggestion. As also suggested by Reviewer #2, we have revised the methods section to present formula to help clarify our model structure, as follows –

"Briefly, for each stratum of country income, a Bayesian model was used for data synthesis of the proportion of RSV hospitalisations by each month of age in the first year of life (the first month was defined as 0-<28 days), as shown in the formula below:

$$\vec{Y}_s \mid (\vec{q}, N_s) \sim Multinomial\ (N_s, \vec{q})$$

$$\vec{q} \mid \vec{\delta} \sim Dirichlet(\vec{\delta})$$

where s denotes each study (from the same stratum of country income); $\vec{q}$ denotes the proportion of RSV hospitalisations in each of the 12 months of age, which follows a Dirichlet distribution; $\vec{Y_s}$ and N_s are the RSV hospitalisation counts in each month of age and totally in the first year of life."

"The age-specific proportion (r_j^k) and age-cumulative proportion (r_j^k) of RSV hospitalisations for a given birth month k could be estimated by:

$$r_j^k = q_j \cdot p_{(k+j-2)\ mod\ 12\ +1}$$

$$r_j^k = \sum_{j=1}^{12} q_j \cdot p_{(k+j-2)\ mod\ 12\ +1}$$

$$\vec{p}\ |\ \vec{y}\ \sim\ Dirichlet(\vec{y}\ +1)$$

where q_j is the posterior estimate of the age-specific proportion of RSV hospitalisation in this study; $\vec{p}$ denotes the proportion of RSV cases in each calendar month of a year, which is simulated from a Dirichlet-Multinomial conjugate model, with a uniform $Dirichlet(1,1,\dots,1)$ prior to represent no prior preference for any months. The posterior distribution of the proportions is then $Dirichlet(1\ +\ y_1, 1\ +\ y_2, \dots, 1\ +\ y_{12})$, where y_{i} are the observed case counts from external data sources."

Additionally, the R codes used in this study are publicly available for researchers interested in conducting similar analyses, which is now added to the code availability section (https://github.com/Glinnnng/Age_distribution_infants)

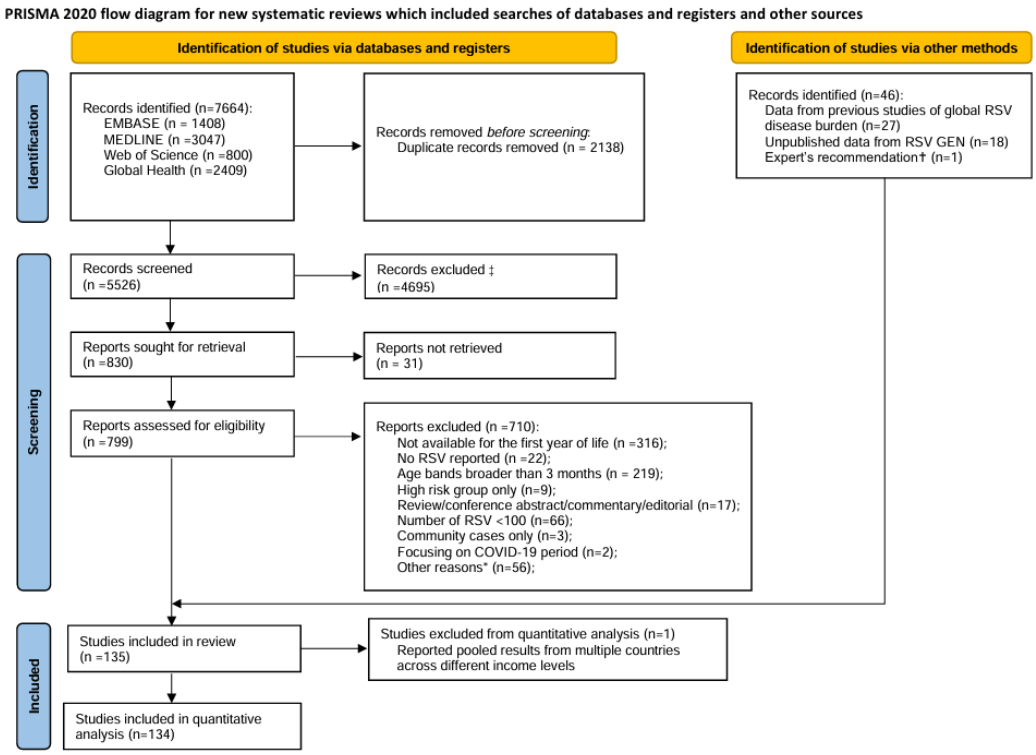
Changes made to the manuscript:

Lines 375-384, Pages 15; Lines 412-422, Pages 16-17; Lines 461-463, Page 18; Supplementary Methods 3.

Comment #2

2. Consider adding a box in Fig 2 to show the study that was excluded from the quantitative analysis because of report of pooled results.

Response: Thanks for the suggestion. Now added (now figure 1).



Changes made to the manuscript:

Figure 1.

Comment #3

3. Could the authors provide a rationale, either in the introduction or discussion, for why they also analyzed data by latitudinal group (temperate, sub-tropical and tropical)?

Response: Thank you for the suggestion. The following text has been added to the method section –

"Given the variations in RSV epidemic patterns at different latitudes[11], as secondary analysis…"

Changes made to the manuscript:

Line 391, Page 16.

Comment #4

4. For the online tool described in lines 258 - 259, I think it would be useful for the authors to allow the user to optionally use percent RSV positivity data as "seasonality data" if the country thinks that case counts could be biased or misleading. Using only cases, as provided for in the prediction tool, requires that testing of patients for RSV is done systematically all through the year. For countries such LICs and LMICs, who could potentially benefit most from this tool, testing for RSV might be conducted differentially for various reasons, including limited testing supplies.

Response: We appreciate the reviewer's insightful suggestion. We agree that in some low- and middle-income countries (LMICs), RSV testing is not done systematically all through the year; as a result, relying solely on case counts may introduce bias. Ideally, the prediction tool should allow for RSV positivity rates as an optional input.

However, it is challenging for us to readily incorporate the option of using percentage RSV positivity data in this study, mainly for the two reasons below:

Limitations of percentage positivity proportion per se

- **Impact of co-circulation of other respiratory viruses**: RSV positivity proportion can be influenced by the co-circulation of other respiratory viruses with similar spectrums of symptoms that prompt testing (e.g., influenza, SARS-CoV-2). When co-circulating with these respiratory viruses, the total number of tests is expected to increase, leading to an artificially lower RSV positivity proportion; by comparison, using the absolute count of RSV would not be affected.

- **Sensitivity to testing methods:** Positivity proportion is highly sensitive to the diagnostic assay used (e.g., PCR, rapid antigen tests), making it difficult to compare across different study locations.

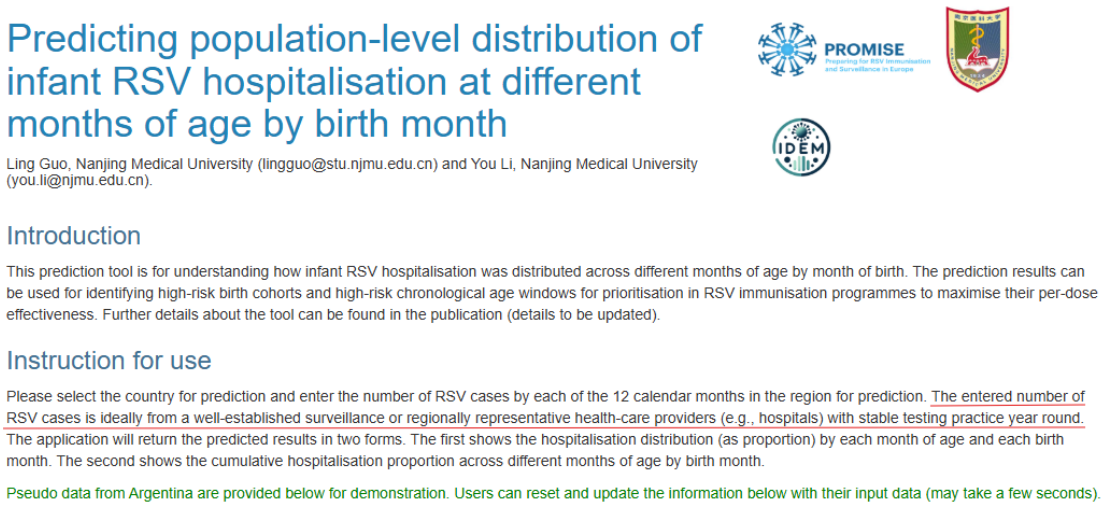
Model Validation Constraints:

- Incorporating the use of RSV positivity requires sufficient data on RSV positivity data for model validation. However, our dataset does not include information on percentage of RSV positivity data required by model validation.

In the revised manuscript, we have discussed the issue of testing stability as follows –

" *Moreover, we assumed that testing practice was stable enough to reflect the true local RSV seasonal patterns; when testing practice varied within a year, further work may consider incorporating other forms of RSV seasonality data such as RSV testing positivity proportion.* "

In addition, we have clarified this at the instruction section on the webpage of the prediction tool –



Changes made to the manuscript:

Lines 287-290, Page 12.

Comment #5

5. In line 293, did the authors mean to say that “a country could include multiple datasets”? It currently reads “study”, yet my understanding is that each dataset is a study.

Response: We specifically chose to use "datasets" rather than "studies" because a single study may report data from multiple countries separately. We have now added an example in the main text to help clarify the term–

“…e.g. a study may provide data for 3 countries at the same time, and the data from each country is considered a separate dataset”

Changes made to the manuscript:

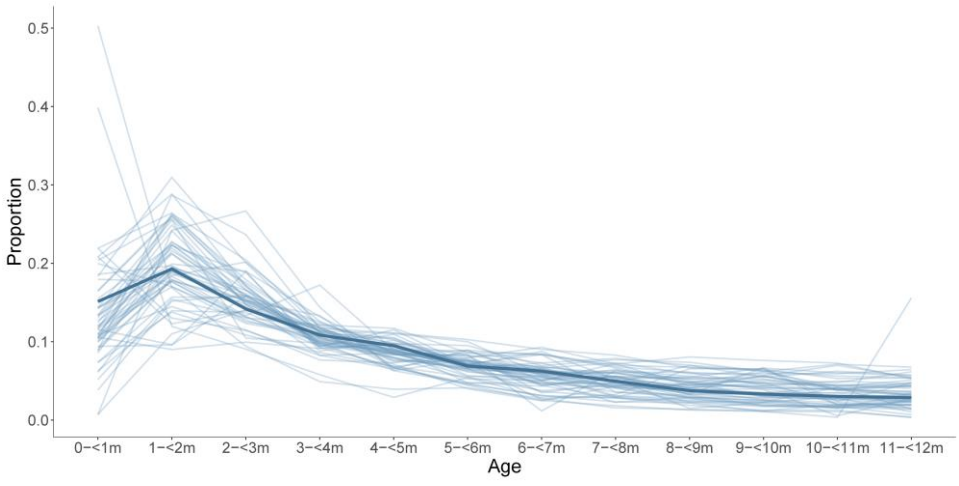
Lines 126-127, Page 6.

Comment #6

6. Might the authors consider showing the data from the ad-hoc verification analysis (described in lines 315 - 320) in a graphical presentation as a supplemental figure similar to Fig 3? This could just be one graph each (for the 134 vs 54 studies) and not necessarily by income status.

Response: Many thanks for the suggestion. Now added in Supplementary Figure 4.

“Supplementary Figure 4. Description of age distribution of RSV hospitalisations in 54 studies that had available data for each month of age.



The light lines indicate the reported proportion of RSV hospitalisations by each individual study, and the dark line denotes the combined proportion of RSV hospitalisations by month of age totalled across individual studies."

Changes made to the manuscript:

Supplementary Figure 4

Reviewer #2 (Remarks to the Author):

Major comments:

Comment #7

1. The authors need to include more details about their model, such as

- What are the observational units? Are they study?
- What exactly are they assuming? Does each study get their own multinomial probability vector that is drawn from a Dirichlet distribution? Or is the probability vector shared across studies? That is, if we let s index study, $\vec{Y}_s$ be a length 12-vector containing the number of hospitalizations at each age in study s, and N_s be the number of hospitalized subjects, is their model

$$\vec{Y}_s \mid (\vec{p}_s, N_s) \sim \text{Multinomial}\ (N_s, \vec{p}_s)$$

$$\vec{p}_s \mid (\alpha, \vec{p}) \sim \text{Dirichlet}(\alpha\ \vec{p}), \alpha > 0, \vec{p} \in [0,1]^{12}, \sum_{m=1}^{12} \vec{p}_m = 1,$$

or is it

$$\vec{Y}_s \mid (\vec{q}, N_s) \sim \text{Multinomial}\ (N_s, \vec{q})$$

$$\vec{q} \mid \vec{\delta} \sim \text{Dirichlet}(\vec{\delta})?$$

The first model is more appropriate, as it captures inter-study variation in hospitalization probabilities. Is this the model they used? If so, what were their priors

on $\vec{p}$ and α? These are needed to implement Metropolis-Hastings.

Response: We agree the need to include more details about the model.

First of all, we would like to clarify the model formula, which is now added to the Methods section of the manuscript as follows –

"Briefly, for each stratum of country income, a Bayesian model was used for data synthesis of the proportion of RSV hospitalisations by each month of age in the first year of life (the first month was defined as 0-<28 days), as shown in the formula below:

$$\vec{Y}_s \mid (\vec{q}, N_s) \sim Multinomial\ (N_s, \vec{q})$$

$$\vec{q} \mid \vec{\delta} \sim Dirichlet(\vec{\delta})$$

where s denotes each study (from the same stratum of country income); $\vec{q}$ denotes the proportion of RSV hospitalisations in each of the 12 months of age, which follows a Dirichlet distribution; $\vec{Y}_s$ and N_s are the RSV hospitalisation counts in each month of age and totally in the first year of life."

Therefore, to answer the questions of the reviewer – the observational units are individual studies; rather than accounting for between-study variations in the model (referred to as the first model by the reviewer), we assume that between-study variations are due to country income strata and therefore use separate models by country income strata.

Changes made to the manuscript:

Lines 375-384, Pages 15.

Comment #8

2. Since the authors did not account for birth month in either the above models, which is very predictive of hospitalization rate, observations within a study are correlated. For example, two subjects born in the same birth month are more likely to have the same hospitalization statuses (hospitalized or not hospitalized) than subjects born in different birth

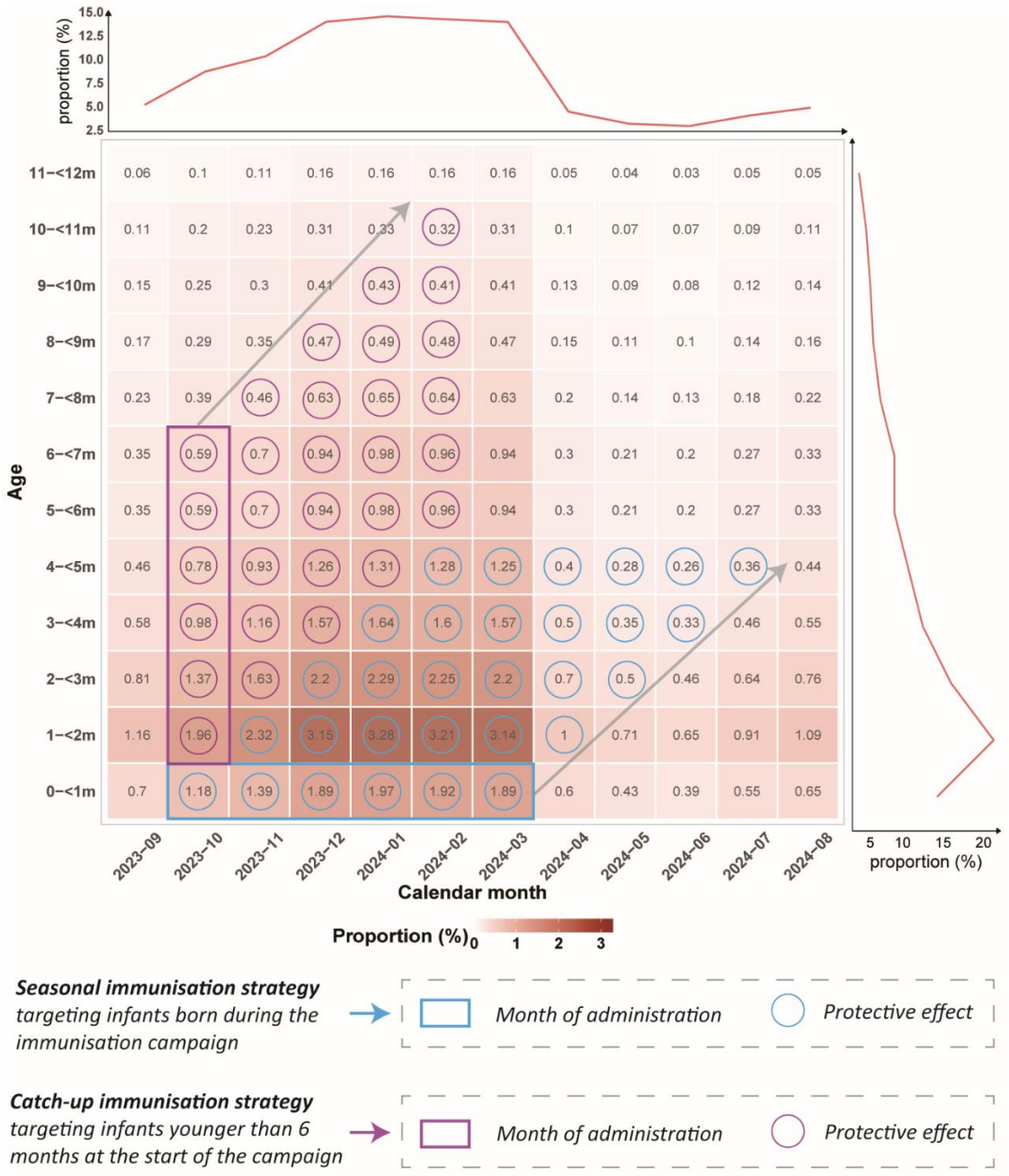
months. As such, the author's multinomial assumption (see Models (1) and (2)), which implicitly assumes hospitalization statuses across subjects are independent, is incorrect. This will cause them to underestimate the width of credible intervals (e.g., Figure 3). If the authors would like to report credible or confidence intervals, they should revise their main model to also account for birth month and other important covariates.

Response: We appreciate the comment and would like to clarify regarding the objective and assumption of the model as well as the assumption regarding the relationship between age, birth month and seasonality.

What we wanted to model first in this study was how the RSV hospitalised cases were distributed among different months of age at the population level, which was further stratified by birth months (in a subsequent separate prediction model), rather than the probability of being hospitalised at the individual level.

The reason for focusing on the distribution (i.e., proportion) rather than probability of hospitalisation (i.e., hospitalisation vs non-hospitalisation) was to address the need for RSV passive immunisation strategy — current recommendation of RSV immunisation for the first year of life is based on whether infants are born during RSV season (i.e., the birth dose) or whether infants are relatively young when experiencing their first RSV season (i.e., the catch-up dose), regardless of perceived individual-level risk factors such as prematurity and comorbidities (which is more relevant to the recommendation for the second year of life, beyond the scope of this study).

To help clarify this, we have added a schematic figure (**Supplementary Methods 4**; see the next page) showing the utility of the model, as shown on the next page. Our model could help identify the optimal immunisation strategy, which should ideally cover the highest proportion of RSV hospitalisations while avoiding covering certain combinations of age and season with limited RSV hospitalisations. We have also made revisions throughout the manuscript to help clarify that the model focuses on the population-level distribution.



(Supplementary Methods 4)

Regarding the assumption of independence, we have provided further clarifications, as follows (in Methods) –

"This was based on the assumption that RSV hospitalisation distribution by birth month was determined by two independent factors, namely chronological age (an inherent factor indicating susceptibility to RSV disease) and local RSV seasonality (an external factor indicating level of exposure to RSV)"

As a child experiences varying levels of RSV circulation at varying age, RSV hospitalisation distribution by birth month is jointly determined by local RSV seasonality and chronological month of age; under the assumption of independence, it will be the product of the proportion of RSV hospitalisations by chronological month of age (based on the estimates of RSV hospitalisation distribution by age in the main analysis) and the proportion of RSV by calendar month of the year be provided externally for the region of interest.

We have added a schematic figure presenting this assumption and how the RSV distribution by birth month was calculated in a new figure (**Figure 4**). We have also provided the full formula for predicting the RSV hospitalisation by birth in the methods section, as follows –

"The age-specific proportion (r_j^k) and age-cumulative proportion (r_j^k) of RSV hospitalisations for a given birth month k could be estimated by:

$$r_j^k = q_j \cdot p_{(k+j-2)\ mod\ 12\ +1}$$

$$r_j^k = \sum_{j=1}^{12} q_j \cdot p_{(k+j-2)\ mod\ 12\ +1}$$

$$\vec{p}\ |\ \vec{y}\ \sim\ Dirichlet(\vec{y}\ +1)$$

where q_j is the posterior estimate of the age-specific proportion of RSV hospitalisation in this study; $\vec{p}$ denotes the proportion of RSV cases in each calendar month of a year, which is simulated from a Dirichlet-Multinomial conjugate model, with a uniform $Dirichlet(1,1,\dots,1)$ prior to represent no prior preference for any months. The posterior distribution of the proportions is then $Dirichlet(1\ +\ y_1, 1\ +\ y_2, \dots, 1\ +\ y_{12})$, where y_{i} are the observed case counts from external data sources."

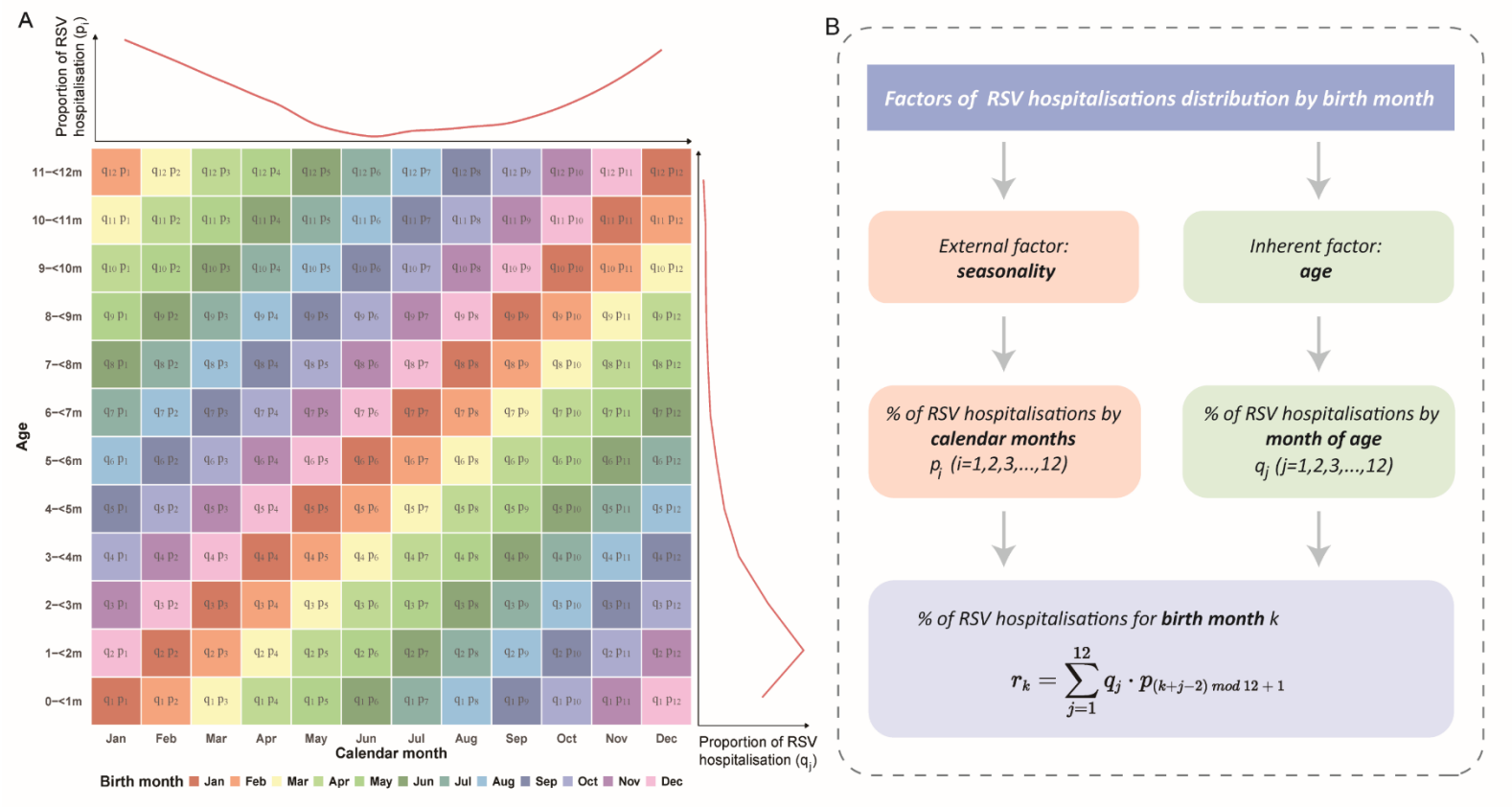


Figure 4. Schematic figure showing the estimation of RSV hospitalisation distribution by birth month.

Panel A shows an example of the relationship between distribution of RSV hospitalisation by chronological month of age and month of birth. Panel B explains how the RSV hospitalisation distribution by birth month is estimated in this study.

Changes made to the manuscript:

Lines 169-177, Page 8; Lines 403-407, Lines 412-422, Pages 16-17; Figure 4; Supplementary Methods 4.

Comment #9

3. I am confused by the assumption required for their online predictor, i.e., "the assumption that the age distribution of RSV cases was independent of any calendar month of the year" (line 241-242). Are they assuming that whether someone is hospitalized with RSV does not depend on the calendar month? If so, this is obviously incorrect due to the seasonality of RSV.

Response: We appreciate the need for further clarifying the assumption of independence. As mentioned in ***Comment #8***, our assumption was that RSV hospitalisation distribution by birth month was determined by two independent factors, namely chronological age (an inherent factor indicating susceptibility to RSV disease) and local RSV seasonality (an external factor indicating level of exposure to RSV). RSV hospitalisation distribution by birth month is therefore jointly determined (independently) by age and by RSV seasonality. We hope the revised descriptions in the manuscript as mentioned in ***Comment #8*** provided further clarifications of the assumption.

Comment #10

4. What is the purpose of the online prediction tool? According to lines 238-241, the authors seem to be estimating the probability a subject is hospitalized at month a given their birth month month B AND THEY ARE EVENTUALLY HOSPITALIZED. As most infants are not hospitalized for RSV, a much more useful tool would be the probability they are hospitalized at month a given their birth month B.

Response: We appreciate comment and realise that there is a need for clarifying the utility

of the prediction tool.

As mentioned in ***Comment #8***, current recommendation of RSV immunisation for the first year of life is based on whether infants are born during RSV season (i.e., the birth dose) or whether infants are relatively young when experiencing their first RSV season (i.e., the catch-up dose), regardless of perceived individual-level risk factors such as prematurity and comorbidities. Therefore, compared with modelling the individual-level risks for RSV hospitalisation (i.e., the individual probability of RSV hospitalisation), understanding how RSV hospitalisations were distributed across different chronological age at different months of the year is more relevant to refining infant RSV immunisation strategy (**Supplementary Methods 4**).

Comment #11

5. The way the authors have presented their online prediction tool is incredibly confusing. They should explicitly state what it is they are trying to calculate, e.g., "the probability an infant is hospitalized at month a given they are born in month B". They should also provide formulas for their calculation. This will make it easier to communicate assumptions.

Response: While we observed mixed comments from reviewers (Reviewer #2 and Reviewer #3) regarding the clarity of the prediction tool, we agreed that further revisions were required.

Following the suggestion, we have revised the introduction section of the tool's webpage as follows –

"This prediction tool is for understanding how infant RSV hospitalisation was distributed across different months of age by month of birth. The prediction results can be used for identifying high-risk birth cohorts and high-risk chronological age windows for prioritisation in RSV immunisation programmes to maximise their per-dose effectiveness."

The formula for the calculation of the prediction tool is now added to the methods section,

as follows –

"The age-specific proportion (r_j^k) and age-cumulative proportion (r_j^k) of RSV hospitalisations for a given birth month k could be estimated by:

$$r_j^k = q_j \cdot p_{(k+j-2)\ mod\ 12\ +1}$$

$$r_j^k = \sum_{j=1}^{12} q_j \cdot p_{(k+j-2)\ mod\ 12\ +1}$$

$$\vec{p}\ |\ \vec{y}\ \sim\ Dirichlet(\vec{y}\ +1)$$

where q_j is the posterior estimate of the age-specific proportion of RSV hospitalisation in this study; $\vec{p}$ denotes the proportion of RSV cases in each calendar month of a year, which is simulated from a Dirichlet-Multinomial conjugate model, with a uniform $Dirichlet(1,1,\ldots,1)$ prior to represent no prior preference for any months. The posterior distribution of the proportions is then $Dirichlet(1\ +\ y_1, 1\ +\ y_2, \ldots, 1\ +\ y_{12})$, where y_{i} are the observed case counts from external data sources."

In the webpage of the prediction tool, we will cite this manuscript following publication.

Changes made to the manuscript:

Lines 412-422, Pages 16-17.

Comment #12

6. Can the authors draw the lines y = x in the top row of plots in Figure 5? While the model predicted and actual probabilities are correlated, it seems that the y-intercepts and slopes are far from 0 and 1, respectively, suggesting the model is poorly calibrated. Can the authors explain why this is happening?

Response: We appreciate the observation and have added the line y = x to the figure as

suggested.

We agreed that the model prediction (compared to the real-world observations) was not optimal, particularly impacted by outliers at both ends. Therefore, we have added some discussion on the reason as well as how it might affect model's application, as follows –

"Fourth, while the external validation results showed that the model had overall consistent distribution of RSV hospitalisations by birth month with the observed data, discrepancy remained in the absolute proportion estimates for some of the birth months, which might potentially stem from variations in the number of births across different calendar months that were not accounted for in the model due to data scarcity. Nonetheless, such discrepancy did not compromise the potential public health implication of the model, which essentially relied on identifying birth cohorts with higher RSV hospitalisation proportion (rather than the precise absolute proportion)."

Changes made to the manuscript:

Lines 290-297, Page 12.

Reviewer #2 (Remarks on code availability):

Comment #13

As far as I can tell, the authors have not made any of their code available.

Response: While we did make our codes available in GitHub when submitted the manuscript, we realised that this information was only available in the Reporting Summary (probably missed by the reviewer) and not in the main manuscript. We have now added the code availability section to the main text, as follows –

"The R codes used in this analysis are available on GitHub (https://github.com/Glinnnng/Age_distribution_infants)"

Changes made to the manuscript:

Lines 461-463, Page 18.

Reviewer #3 (Remarks to the Author):

This is an elegant study determining the risk of respiratory syncytial virus (RSV) hospitalization using two important and readily available risk factors: chronological age and birth month. The authors developed a country/latitude specific prediction tool predicting country specific RSV hospitalization burden. The only information that needs to feed into the prediction tool is monthly RSV surveillance data. The prediction tool can be of significant public health importance, guiding country/region specific RSV immunization programs and improving vaccine efficiency. My main comments are:

Comment #14

1. Validity of the underlying assumption of independence of chronological age and birth month on the risk of RSV hospitalization. The prediction tool is developed with this underlying assumption. Throughout the manuscript, the relationship between these two factors and the risk of infant RSV hospitalization is presented separately, though at the same time, interpreted as there is an interaction effect. It is confusing whether the effect of the two factors is truly independent or not. The varying cumulative risk in infancy due to birth month hints an interaction effect (supplemental figure).

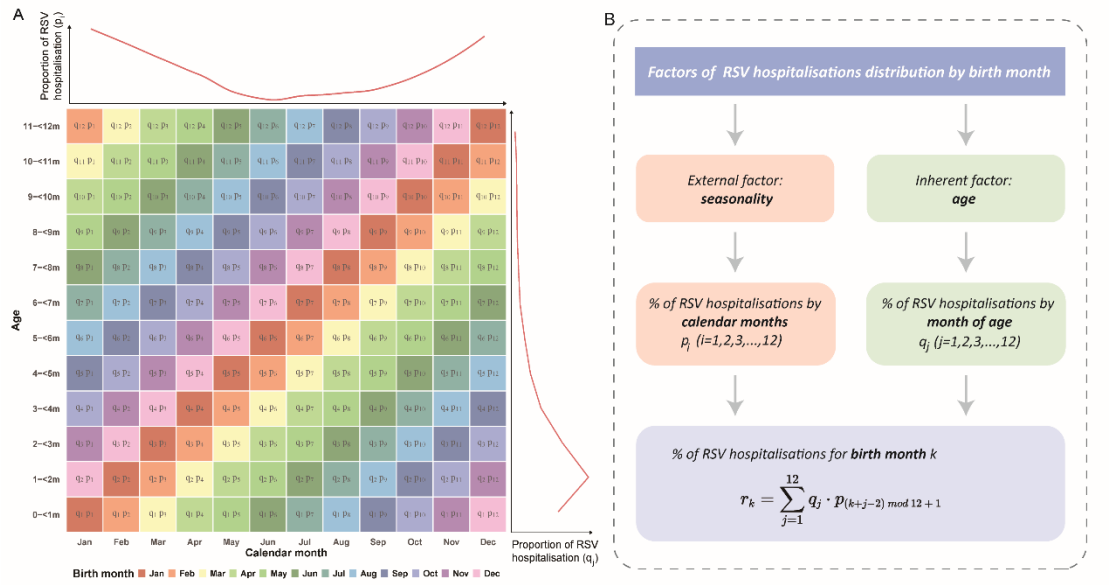
Response: We appreciate the comment. We would like to clarify that our model assumed independence between chronological age (an inherent factor for RSV hospitalisation determining susceptibility) and seasonality (an external factor for RSV hospitalisation determining risks of exposure), rather than between age and birth month. In fact, the distribution of RSV hospitalisation by birth month is jointly determined by the two factors above and reflects the interaction, precisely as the reviewer has stated.

We have now expanded both the Results and Methods sections to help clarify the assumption of independence.

Results

"For developing the prediction tool for RSV hospitalisation by birth month, we considered two key independent determinants at the population level: chronological month of age, which reflects the changes in inherent susceptibility to RSV diseases, and RSV circulation in the community (i.e., seasonality), which is external factor that determines the risk of exposure to RSV. As a child experiences varying levels of RSV circulation at varying age, RSV hospitalisation distribution by birth month is jointly determined by local RSV seasonality and chronological month of age…"

Figure 4–



Methods–

"This was based on the assumption that RSV hospitalisation distribution by birth month was determined by two independent factors, namely chronological age (an inherent factor indicating susceptibility to RSV disease) and local RSV seasonality (an external factor indicating level of exposure to RSV); an illustrative example of this model is presented in Supplementary Methods 4. Under this assumption, the proportion of RSV hospitalisations per birth month was the product of the proportion of RSV hospitalisations by chronological month of age (based on the estimates of RSV hospitalisation distribution by age in the main

analysis) and the proportion of RSV by calendar month of the year to be provided externally for the region of interest (Fig. 4).”

Changes made to the manuscript:

Lines 169-177, Page 8; Lines 403-412, Page 16; Figure 4.

Comment #15

2. The RSV hospitalization burden presented in the manuscript is relative to the overall infancy RSV hospitalization (proportions). Infant RSV hospitalization prevalence itself varies by countries. It might be even more policy impactful to “predict” country specific month-by-month absolute number of RSV hospitalizations and the associated costs for each birth month.

Response: We greatly appreciate the comment. We agree that in addition to predict the relative hospitalisation (to the overall infant RSV hospitalisation), it would be even more policy impactful for predicting the absolute RSV hospitalisation burden. Predicting county-level RSV hospitalisation burden required a further understanding of determinants and available data on these determinants. However, we did not have data readily available for expanding the current model framework for such predictions, which could be possible as a separate study.

Comment #16

3. The study focuses on RSV hospitalization; it is important to note that RSV hospitalization is only a small fraction of infant medically attended RSV illnesses. Infant RSV associated disease burden will be higher

Response: We appreciate your comment. We recognise that our current analysis focused specifically on the distribution of RSV hospitalisations, while other critical aspects such as

community-based cases were not assessed. We have now discussed this in as a limitation as follows –

"Moreover, we considered only RSV hospitalisation in this study, the more severe layer of the RSV disease spectrum, while RSV disease burden in the community and primary-care settings can be also important for decision-making in the implementation of RSV immunisations."

Changes made to the manuscript:

Lines 297-300, Page 12.

Comment #17

4. Limiting the risk factor to chronological age and birth month is the strength of the study, it also limits the application of the study/prediction tool to more public health policy level as many subject specific risk factors such as preterm birth, Down syndrome, chronic lung disease due to prematurity… are not included in the prediction tool.

Response: We appreciate the comment regarding the inclusion of individual-level risk factors such as preterm birth, Down's syndrome, chronic lung diseases, etc.

Currently, infant RSV passive immunisation recommendations follow a seasonal approach for the general population, i.e., regardless of individual risk status (the high-risk approach is current relevant to children at the second year of life). The prediction tool developed in this study is specifically designed to optimise RSV infant immunisation strategy, maintaining as simple as possible to minimise data requirement. Therefore, we have decided to not to pursue prediction with individual-level risk factors.

We have now clarified this in the discussion section as follows –

"As current recommendation of RSV immunisation for the first year of life is based on whether infants are born during RSV season (i.e., the birth dose) or whether infants are relatively young when experiencing their first RSV season (i.e., the catch-up dose), regardless

of perceived individual-level risk factors such as prematurity and comorbidities (which is more relevant to the recommendation for the second year of life, beyond the scope of this study). our study focused only on the overall distribution of RSV hospitalisation burden at the general population level, and did not intend to predict individual risks for RSV hospitalisation that could also be driven by individual-level risk factors."

Changes made to the manuscript:

Lines 300-308, Page 12.

Reviewer #3 (Remarks on code availability):

Comment #18

The web link of the prediction tool is very user friendly and self explanatory

Response: Thank you for the positive feedback.

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

Comment #1

Thanks for your response. My comments have sufficiently been addressed.

Response: Thank you.

Reviewer #2 (Remarks to the Author):

Comment #2

1) Response to my first comment: in my opinion, the model the authors are using to model the number of hospitalizations is not appropriate. According to their response, they are assuming that all observations within the same income stratum are independent. This is not realistic, as it is much more likely that observations from the same study are more similar to one another than observations from different studies. Put another way, observations from the same study are correlated. While this issue of correlation will likely not have an impact on point estimates (e.g., their solid lines in Figure 2), the credible intervals they are reporting are likely much too narrow. The authors can account for within-study dependence using the first model I proposed in Comment 1 of my original response. They should also be accounting for birth month if they would like credible intervals to be believable (see below).

Response: Thank you. We appreciate the suggestion and the detailed explanation. We have now adopted the model as per suggestion. Here, we chose to estimate directly the between-study variations by conducting a separate Dirichlet regression using a subset of our dataset and estimating the concentration parameter that quantifies the variations, which should have already included the variations

associated with birth month as a single factor. We have updated the methods section as follows –

"Briefly, for each stratum of country income, a hierarchical Bayesian model was used for data synthesis of the proportion of RSV hospitalisations by each month of age in the first year of life (the first month was defined as 0-<28 days) while accounting for between-study variations, as shown in the formula below:

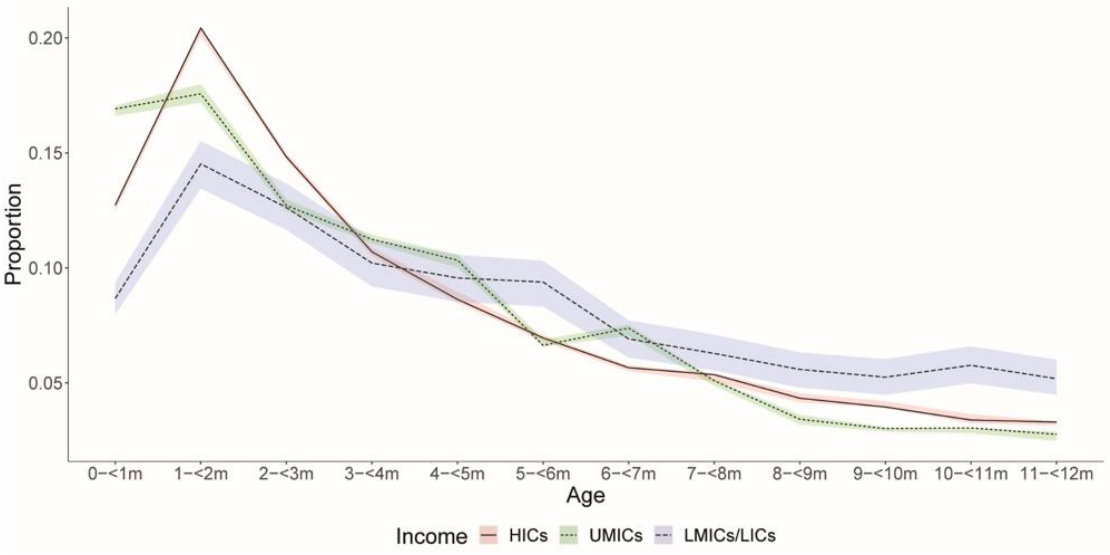
$$\vec{Y}_s \mid (\vec{p}_s, N_s) \sim \text{Multinomial}\ (N_s, \vec{p}_s)$$

$$\vec{p}_s \mid (\alpha, \vec{p}) \sim \text{Dirichlet}(\alpha\ \vec{p}), \alpha > 0, \vec{p} \in [0,1]^{12}, \sum_{m=1}^{12} \vec{p}_m = 1$$

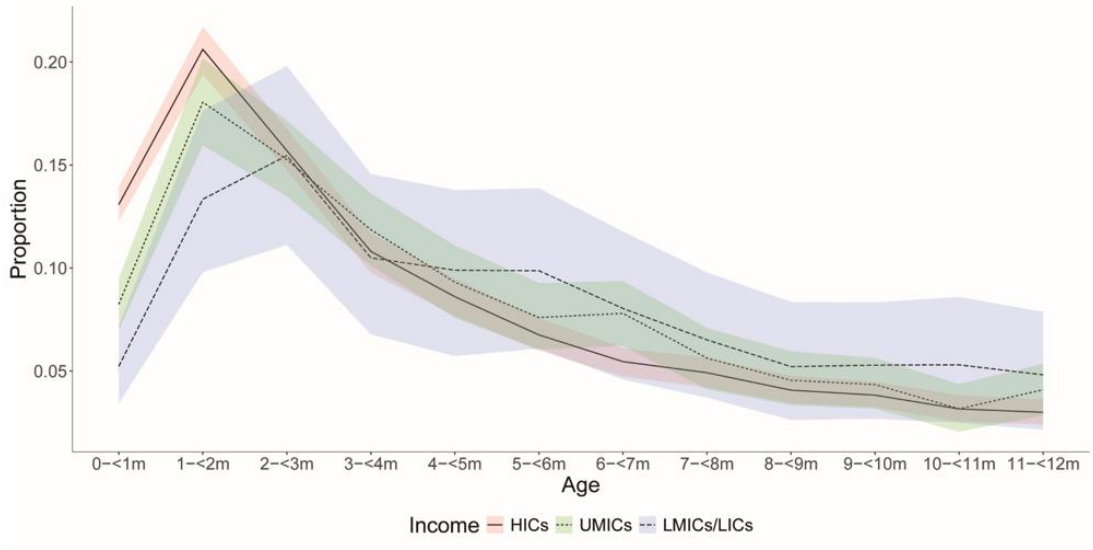
where s denotes each study (from the same stratum of country income); $\vec{p}_s$ denotes the proportion of RSV hospitalisations in each of the 12 months of age for each study, which follows a Dirichlet distribution; $\vec{Y}_s$ and N_s are the RSV hospitalisation counts in each month of age and totally in the first year of life; α, as the concentration parameter of the Dirichlet distribution, represents the heterogeneity among studies within the same country income stratum. Here, α was set to a fixed value of 100, which was informed by a separate Dirichlet regression of a subset of our dataset that had available RSV hospitalisation counts for each month of age. $\vec{p}$ denotes the proportion of RSV hospitalisations in each of the 12 months of age for each stratum of country income, which was the primary parameter of interest."

Regarding the results, as expected, the point estimate was generally consistent to the previous model (though it was more variable in low- and lower-middle-income countries due to smaller sample size) and it is only the width of the credible intervals that has been changed. We have presented here the original Figure 2 (panel A) and updated Figure 2 (panel A) as an example –

The original



The updated



With the major revision to the model, we believe the updated estimates and their credible intervals are more reasonable. We have updated the results section and figures throughout the manuscript accordingly.

Changes made to the manuscript:

Line 71, Page 4; Lines 136-144, Pages 6-7; Lines 155-161 and 165-166, Page 7; Lines 183 and 188-189, Page 8; Lines 197, 200, 205-207 and 211-214, Page 9; Lines 218-224, Pages 9-10; Lines 232-233, 238-239, 242-243, 246, and 249-250, Page 10; Lines

376-393, Pages 15-16; Lines 398-399, Page 16; Supplementary Methods 3 and 5; Supplementary Figures 3, and 5-13; Supplementary Tables 5-6.

Comment #3

2) Response to my second comment: unfortunately, the authors did not address my concern. I had concerns about the within-study dependence caused by not accounting for birth month in their proposed model, not their stated objectives (which are excellent!) As noted in my second comment, this dependence implies their credible intervals (Figures 2 and 3) are too narrow. If the authors cannot account for birth month in their model (whose estimates are used to generate Figures 2 and 3) or cannot show that birth month has a negligible impact on hospitalizations, I would recommend removing credible intervals entirely or state that they are likely too narrow.

Response: We appreciate the suggestion. As mentioned in the previous response, we have now revised the model to account for the within-study dependence (or in other words, the between-study variations), which we hope has addressed the reviewer's concern here.

Comment #4

Reviewer #2 (Remarks on code availability):

The code looks great.

Response: Thank you.

Comment #5

Reviewer #3 (Remarks to the Author):

All my comments are well addressed.

Response: Thank you.

Major comments:

1. The authors need to include more details about their model, such as

   - What are the observational units? Are they study?
   - What exactly are they assuming? Does each study get their own multinomial probability vector that is drawn from a Dirichlet distribution? Or is the probability vector shared across studies? That is, if we let s index study, $\vec{Y}_s$ be a length 12-vector containing the number of hospitalizations at each age in study s, and N_s be the number of hospitalized subjects, is their model

   $$\begin{aligned} \vec{Y}_s \mid (\vec{p}_s, N_s) &\sim \text{Multinomial}(N_s, \vec{p}_s) \\ \vec{p}_s \mid (\alpha, \vec{p}) &\sim \text{Dirichlet}(\alpha\vec{p}), \quad \alpha > 0,\ \vec{p} \in [0,1]^{12},\ \sum_{m=1}^{12} \vec{p}_m = 1, \end{aligned} \tag{1}$$

   or is it

   $$\begin{aligned} \vec{Y}_s \mid (\vec{q}, N_s) &\sim \text{Multinomial}(N_s, \vec{q}) \\ \vec{q} \mid \vec{\delta} &\sim \text{Dirichlet}(\vec{\delta})? \end{aligned} \tag{2}$$

   The first model is more appropriate, as it captures inter-study variation in hospitalization probabilities. Is this the model they used? If so, what were their priors on $\vec{p}$ and α? These are needed to implement Metropolis-Hastings.

2. Since the authors did not account for birth month in either the above models, which is very predictive of hospitalization rate, observations within a study are correlated. For example, two subjects born in the same birth month are more likely to have the same hospitalization statuses (hospitalized or not hospitalized) than subjects born in different birth months. As such, the author's multinomial assumption (see Models (1) and (2)), which implicitly assumes hospitalization statuses across subjects are independent, is incorrect. This will cause them to underestimate the width of credible intervals (e.g., Figure 3). If the authors would like to report credible or confidence intervals, they should revise their main model to also account for birth month and other important covariates.

3. I am confused by the assumption required for their online predictor, i.e., "the assumption that the age distribution of RSV cases was independent of any calendar month of the year" (line 241-242). Are they assuming that whether someone is hospitalized with RSV does not depend on the calendar month? If so, this is obviously incorrect due to the seasonality of RSV.

4. What is the purpose of the online prediction tool? According to lines 238-241, the authors seem to be estimating the probability a subject is hospitalized at month a given their birth month month B AND THEY ARE EVENTUALLY HOSPITALIZED. As most infants are not hospitalized for RSV, a much more useful tool would be the probability they are hospitalized at month a given their birth month B.

5. The way the authors have presented their online prediction tool is incredibly confusing. They should explicitly state what it is they are trying to calculate, e.g., "the probability an infant is hospitalized at month a given they are born in month B". They should also provide formulas for their calculation. This will make it easier to communicate assumptions.

6. Can the authors draw the lines $y = x$ in the top row of plots in Figure 5? While the model predicted and actual probabilities are correlated, it seems that the y-intercepts and slopes are far from 0 and 1, respectively, suggesting the model is poorly calibrated. Can the authors explain why this is happening?