# Global, Regional and National Estimation of Firstborn Prevalence, 1990-2022

Kathryn M. Lau

A thesis
submitted in partial fulfillment of the
requirements for the degree of

Master of Public Health

University of Washington

2020

Committee:

Nicholas Kassebaum

Haidong Wang

Program Authorized to Offer Degree:
Global Health

University of Washington

**Abstract**

Global, Regional and National Estimation of Firstborn Prevalence, 1990-2022

Kathryn M. Lau

Chair of the Supervisory Committee:
Nicholas Kassebaum
Department of Global Health

*Introduction:* Birth order, and in particular whether or not a child is firstborn, is known to be directly correlated with risk of specific neonatal diseases and indirectly associated with health outcomes beyond the neonatal period. Accurate and detailed knowledge of firstborn prevalence is therefore important in its own right and also a powerful tool for improving our understanding of multiple child health conditions. Data on birth order is not available in every location or year, so modeling is required. Previous models have used all age total fertility rate as a covariate which, although simpler, obscures important cohort fertility trends and subsequently may bias results. This analysis develops a new modeling strategy for estimating firstborn prevalence that explicitly incorporates the interaction between maternal age and historical fertility trends of maternal cohorts, exploring the model and prediction performance comparisons between the two approaches.

*Methods:* Data on birth order from population-representative surveys and vital registration statistics were extracted and tabulated to inform an annual model of firstborn prevalence. Total fertility rate estimates from the Institute for Health Metrics and Evaluation were recomputed to derive a measure of fertility specific to maternal age and cohort, and the new age-specific measure was used as a covariate in a three step spatiotemporal Gaussian process regression to estimate maternal-age-specific firstborn prevalence across all GBD 2020 locations annually from 1990 to 2022. To first directly compare the effect of modeling by maternal age compared to across all ages, one dataset restricted to sources with complete coverage of the reproductive age range was modeled using both approaches and assessed through in-sample error metrics and qualitative comparison. To generate more comprehensive final estimates, a third model was then run using the age-specific model settings and full dataset (no restrictions on complete age coverage).

*Results:* When run on the same data, the age-specific model performed better than the all age model in terms of in-sample fit. Both models had a small negative bias (-3.55% in the all-age model and -2.85% in the age-specific model). When the age-specific model was run on the full dataset, firstborn prevalence globally in 2019 was estimated to be 38.9% (36.9 – 40.9), an increase of 8.72% (1.50 – 16.8) since 1990. Firstborn prevalence in 2019 was highest in the Republic of Korea and Taiwan and lowest in Guinea-Bissau and the Solomon Islands. In 1990 firstborn prevalence was highest in San Marino and Greenland and lowest in Tuvalu and Qatar. Globally, firstborn prevalence was highest among women 10-14 years old and lowest among women 45-49 years old.

*Discussion:* Modeling firstborn prevalence by maternal age improves upon modeling across all ages by improving in-sample fit, removing assumptions about comparability of maternal age ranges, and producing age-specific estimates that are more informative and more actionable. The set of complete and cross-nationally comparable estimates of firstborn prevalence can be used as predictive covariates to model other health outcomes, and the age-specific modeling framework reveals more specific drivers of changes in firstborn prevalence and presents opportunities for modeling additional fertility scenarios.

## Introduction

Birth order, defined in demography as the ordinal rank of a live birth in a woman's lifetime, has been hypothesized to influence a wide range of health outcomes, including infant mortality, vaccination rates, and long-term mental health.[1–4] While many of these associations require more research to establish a mechanism of action, one outcome with a well-understood relationship with birth order is prevalence of neonatal hemolytic disease due to rhesus (Rh) alloimmunization. Firstborns (newborns with a birth order of one) are not at risk of this disease, whereas any newborn with a later birth order is potentially at risk. Hemolytic disease due to Rh alloimmunization develops when a mother with an Rh negative blood type gives birth to an Rh-positive newborn. The mismatch in blood types triggers a maternal immune response, which does not harm the firstborn child, but places all subsequent children at risk of hemolytic disease triggered by maternal antibodies against the Rh antigens on the newborn's Rh-positive red blood cells.[5] Accurate estimates of firstborn prevalence are therefore highly useful in etiologic models of hemolytic disease burden, and they are used, for example, in the modeling of neonatal hemolytic disease in the Global Burden of Disease project.[6]

When data on birth order is routinely and reliably collected, firstborn prevalence can be tabulated directly. However many countries do not have complete vital registration systems or do not include birth order.[7] Therefore models are necessary to predict firstborn prevalence in the absence of complete data. To address data gaps and produce a complete time-series of firstborn prevalence that is also comparable across geographies, this analysis seeks to use high-quality data on birth order in location-years where it can be found and relevant fertility covariates to estimate firstborn prevalence in all location-years through a predictive modeling approach. Complete estimates of fertility are available from research groups like the Institute for Health Metrics and Evaluation,[8] and fertility is an intuitive choice as a covariate for firstborn prevalence because the number of children a woman gives birth to is inherently related to the distribution of firstborn versus later born children.

One current method of estimating firstborn prevalence utilizes total fertility rate (TFR) as a predictor.[6] TFR represents the expected number of live births per woman if they passed through the reproductive years bearing children at current age-specific fertility rates (ASFR), where ASFR is the annual number of live births per woman of a given age or age group.[9] Predicting firstborn prevalence by year using TFR raises two potential issues. First, TFR summarizes fertility over all

reproductive ages, and therefore does not account for differences in maternal age composition between various countries and years. Second, TFR is based on fertility in a current year and does not reflect the past fertility rates that each maternal cohort has actually experienced. In this analysis, I propose a new, evidenced-based modeling approach to estimate the prevalence of firstborns annually from 1990 to 2022, at the global, regional, national, and selected subnational levels, with a focus on accounting for maternal age composition and age-specific temporal fertility trends.

**Methods**

This analysis consisted of two main aims. The first aim was to directly compare the model performance and results of a model that accounted for maternal age to a model that did not. The second aim was to use the validated age-specific model to estimate firstborn prevalence using all available data. To accomplish the first aim, the two modeling approaches needed to be applied to the same data. Otherwise, differences in model performance and results could be due to compositional bias in what data went into each model. To ensure equivalent data when comparing models, only data sources with full coverage of the reproductive age range of 10-54 years were used because only that data can be accurately expressed as both an all-age aggregate value and as a comprehensive set of age-specific values. For the second aim, the data restriction was not necessary and a broader dataset that included any age-specific data was used.

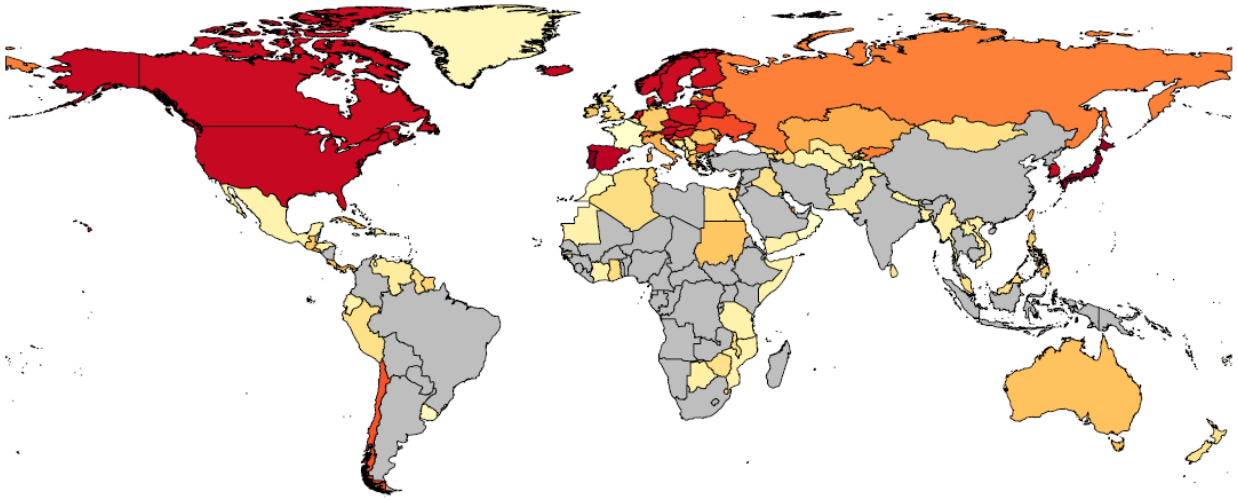
*Data Sources*

Data was included from population-representative surveys and vital registration statistics. Data was identified from three main sources: the United Nations Demographic Yearbook,[10] the Human Fertility Database,[11] and published surveys. The majority of surveys were from the multinational Demographic and Health Survey series,[12] with some additional survey data sources from the Multiple Indicator Cluster Surveys,[13] Reproductive Health Surveys,[14] and country-specific surveys. Only data specifying births by birth order, location, year and maternal age were included. In 314 unique location-years, more than 1% of the total births had birth order labelled as "Unknown" and were excluded. In 239 unique location-years, at least one maternal age group was wider than the standard five year bins used in this analysis (i.e. "40 +"), and data in these wide bins were therefore also excluded. The geographic distribution of the dataset before these exclusions is

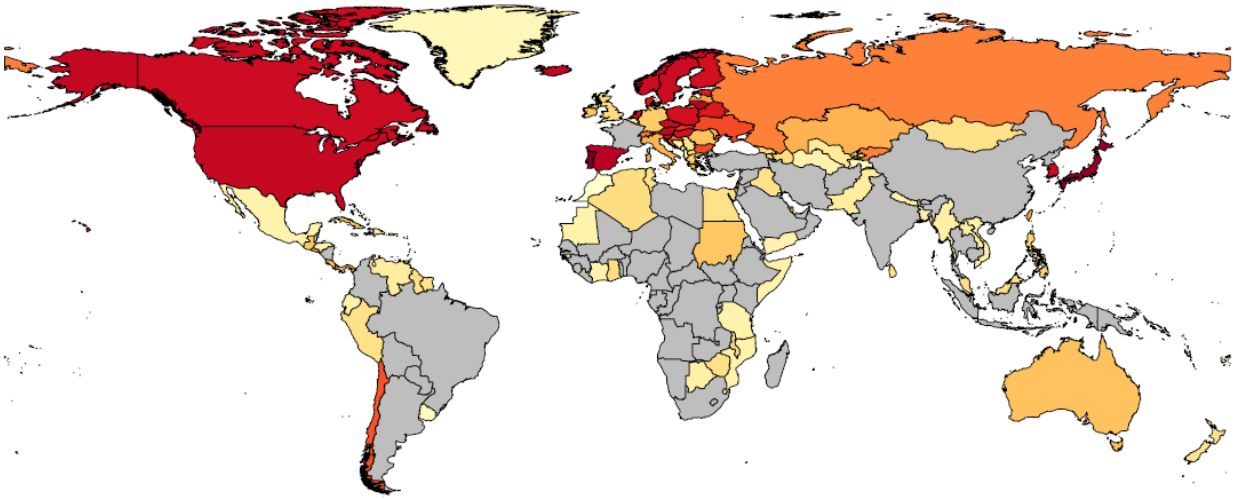
shown in Figure 1A. Figure 1B depicts the coverage of the dataset after exclusion of data where more than 1% of births had unknown birth order or maternal age was provided in age bins wider than five years. This dataset, which included 2,341 location-years of data, covering 126 countries and territories, was used in the final age-specific model. To directly compare the effect of modeling by maternal age versus across all ages, a version restricted to sources with complete coverage of the reproductive age range of 10-54 years was also saved. This version consisted of 1,142 location-years of data, covering 69 countries and territories, and the geographic distribution is shown in Figure 1C.

**Figure 1. Geographic Coverage of Data on Birth Order.**

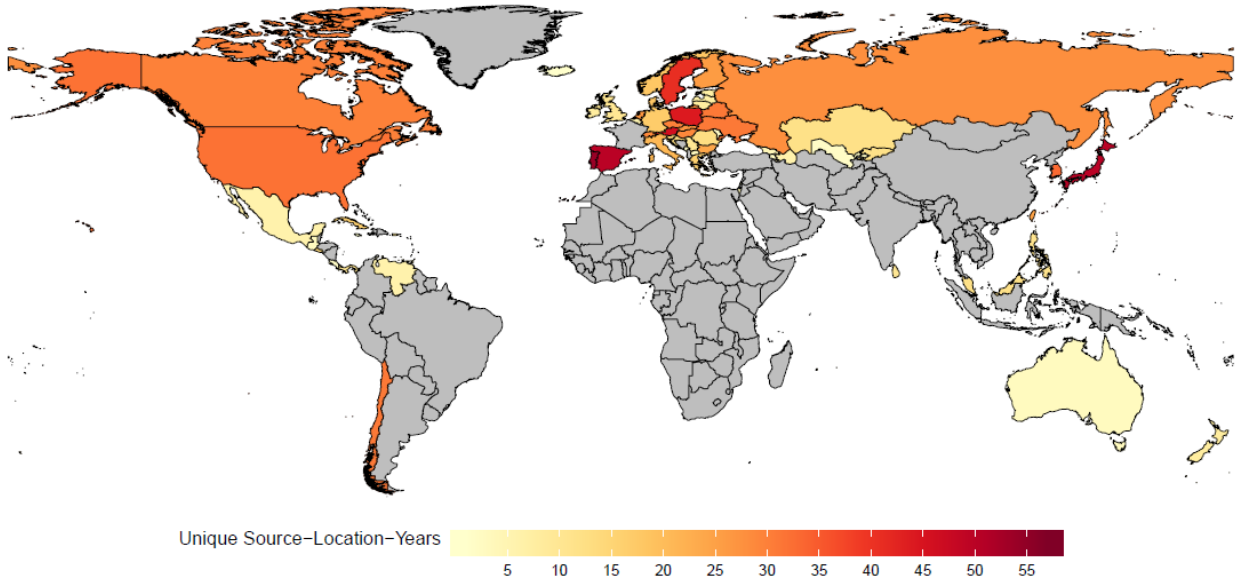
**A. Data with Any Maternal Age Detail and Allowing Unknown Birth Order**



**B. Data with Maternal Age in 5 Year Age Bins and <1% Unknown Birth Order**



**C. Data with Maternal Age in 5 Year Age Bins, <1% Unknown Birth Order, and Complete Age Range Coverage**




*Data Processing*

The primary outcome was firstborn prevalence, calculated as the number of firstborns divided by the number of live births. Numeric birth order data was tabulated, taking the sum of live births of birth order one as the numerator and the sum of live births of any birth order as the denominator. Data were stratified by location, year of birth, and maternal age. Maternal age was defined as the age range from 10 to 54 years and binned into five-year age groups.

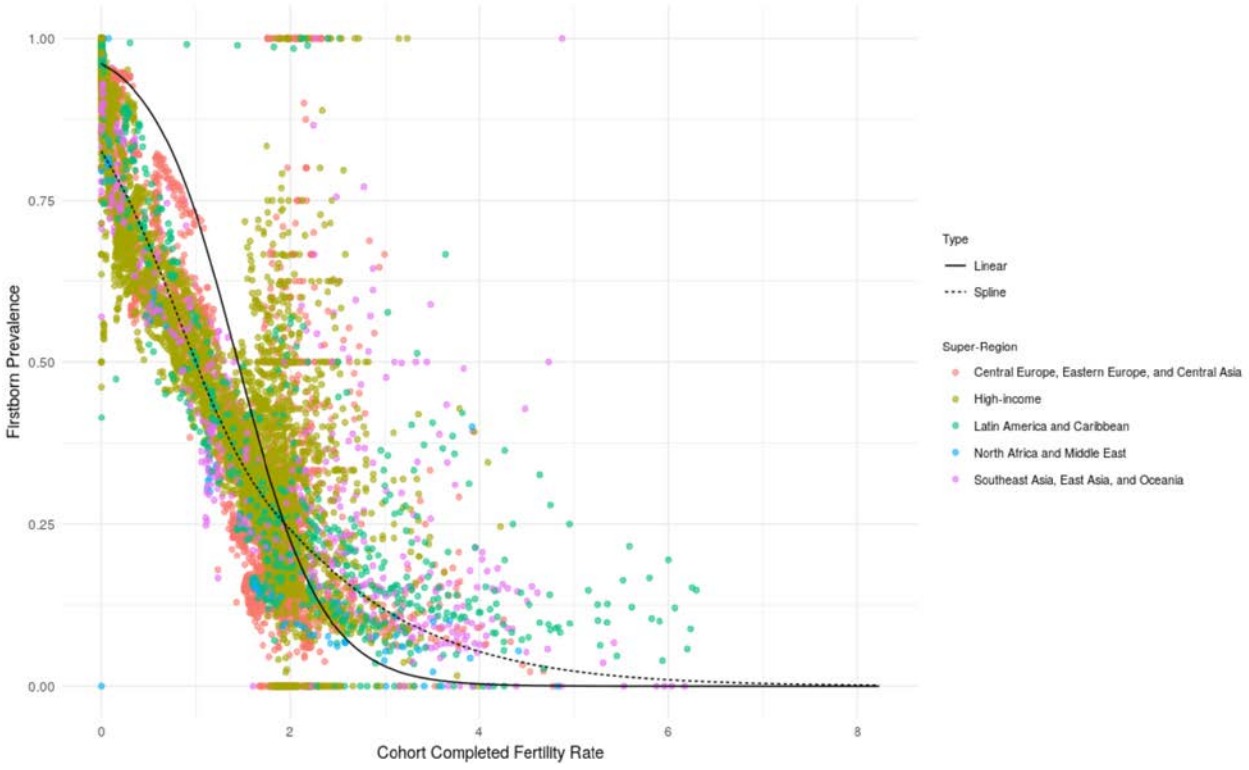
*Covariate Construction*

To account for maternal age composition, a measure of cohort fertility was derived from ASFR estimates produced by IHME.[8] Cohort completed fertility rate (CCFR) was calculated as the sum of ASFRs up to a specified age and year for an individual maternal cohort. For example, the CCFR for 20-24 year olds in 2010 is the sum of the ASFRs experienced by that cohort up to that point in time, namely the ASFR for 20-24 year olds in 2010, 15-19 year olds in 2005, and 10-14 year olds in 2000. IHME's ASFR estimates were available starting from 1950, so given that calculating a full series of CCFR required forty years of ASFR data, the covariate was constructed from 1990 forward. CCFR was calculated for every year, every country or subnational region, and every maternal age group from 1990 to 2022.

*Modeling Strategy*

Firstborn prevalence was modeled using a three-stage spatiotemporal Gaussian process regression (ST-GPR), originally developed by researchers on the Global Burden of Disease study.[15] ST-GPR is a stochastic modeling technique that supports estimation of non-linear trends. The first stage of ST-GPR is to model the general trend based on predictive covariates, typically through a linear model. For this analysis, a linear mixed effects model with a fixed effect on CCFR and random effects on location, and a spline on CCFR with random effects on location were both tested to model the logit-transformed firstborn prevalence. The spline was selected as better capturing the trends in the data, particularly for the extreme values of CCFR (Figure 2). A cubic spline with knots set based on frequency was used (run using a meta-regression package called MR-BRT developed by IHME).[16] The regression used all location-years of available data to generate predictions for every location-year of interest.

**Figure 2. First-stage model of firstborn prevalence based on CCFR: comparison between linear and spline**




In the second stage, spatiotemporal weights were calculated based on the residuals between the first stage estimates and the original data, and the weights were applied to smooth

the residuals across space, time, and age. The degree of smoothing was defined through three hyper-parameters: ζ (space), λ (time) and ω (maternal age). Initial candidate values for the hyper-parameters were selected based on qualitative observations of the degree of correlation across space, time and maternal age in the data. The model was then tested iteratively and the combination of hyper-parameters with the best fit was chosen (Table 1).

**Table 1. Hyper-parameters used to control degree of smoothing**

| Model Hyper-parameter | Value |
|---|---|
| ζ (space) | 0.01 |
| λ (time) | 0.05 |
| ω (maternal age) | 2 |

In the third stage, a Gaussian process regression was run to perform additional smoothing of the residuals between the original data and the second stage results and generate a full time series of results. The third stage also estimated uncertainty of the final results. To estimate uncertainty, 1,000 draws were randomly sampled from a normal distribution of each demographic combination. The final result was the mean of the draws and a 95% uncertainty interval was calculated from the 2.5 and 97.5 percentile of the distribution.

To analyze the effect of modeling with an age-specific measure of fertility, a comparison model was run using the same parameters, but with aggregated data and a spline on all age, non-cohort TFR (the traditional measure). The detailed age-specific data was aggregated into all-age data to ensure the same data was used in each model, and therefore differences in the results could be isolated to differences in covariate and demographic construct of the model. A final model was then run with the same parameters on the full dataset, with no exclusion of data that only partially covered the maternal age range.

*Model Comparison*

The performance of the first two models were compared directly with three quantitative metrics to assess in-sample performance. Using the all-age data and the all-age model results, as well as the all-age data and an aggregation of the age-specific model results, root mean squared error (RMSE) was calculated to assess how well the mean values of the estimates matched the data, and data coverage was calculated to assess how much of the data fell within the estimated uncertainty interval. Mean relative error (MRE) was also computed to identify the bias in each

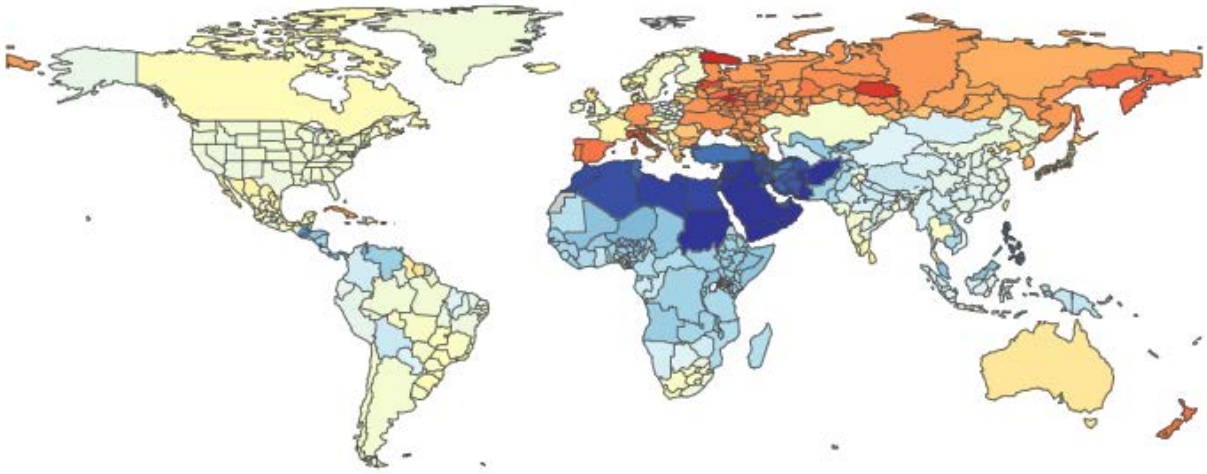
model. In addition, the results of the two model specifications – age/cohort-specific versus all ages – were compared through: 1) comparison of absolute numbers and time trends by country, region, super-region, and globally; 2) qualitative assessment of model fit to data for each specific location, with a special focus on locations and years without input data. This comparison approach of examination of absolute numbers and trends, and qualitative assessment of model fit, was repeated to analyze the impact of switching from the restricted dataset to the full dataset in the age-specific model.

**Results**

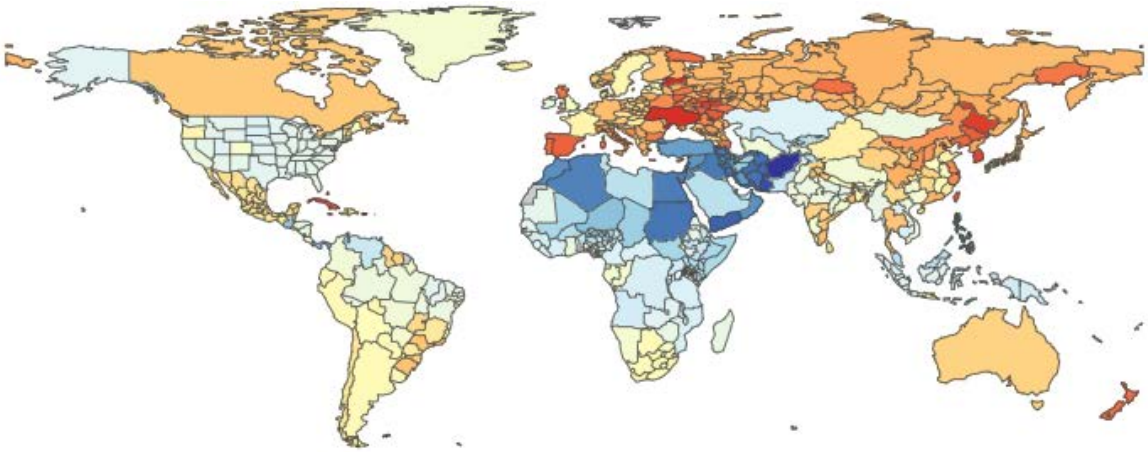
*Model Comparison*

The global estimate of firstborn prevalence in 2019 was 37.9% (95% Uncertainty Interval [UI] 37.8 – 38.1) in the all age model and 35.1% (33.1 – 36.9) in the age-specific model. This was an increase of 4.41% (3.93 – 4.86) since 1990 in the all age model and 12.8% (4.94 – 20.3) since 1990 in the age-specific model. The two maps below show the geographic distribution of firstborn prevalence in 1990 and 2019. Figure 3 presents the results of the all age model and Figure 4 presents the aggregated results of the age-specific model.

**Figure 3. Geographic distribution of firstborn prevalence in 1990 and 2019, according to the all age model.**



A. 1990



B. 2019

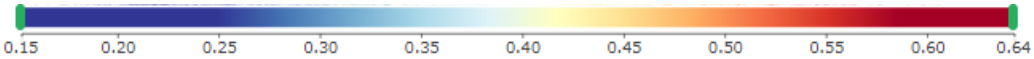
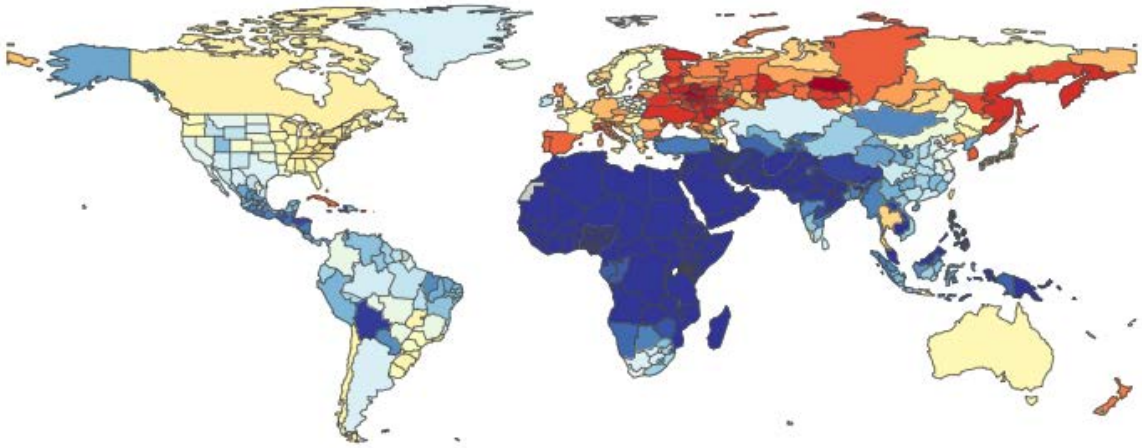
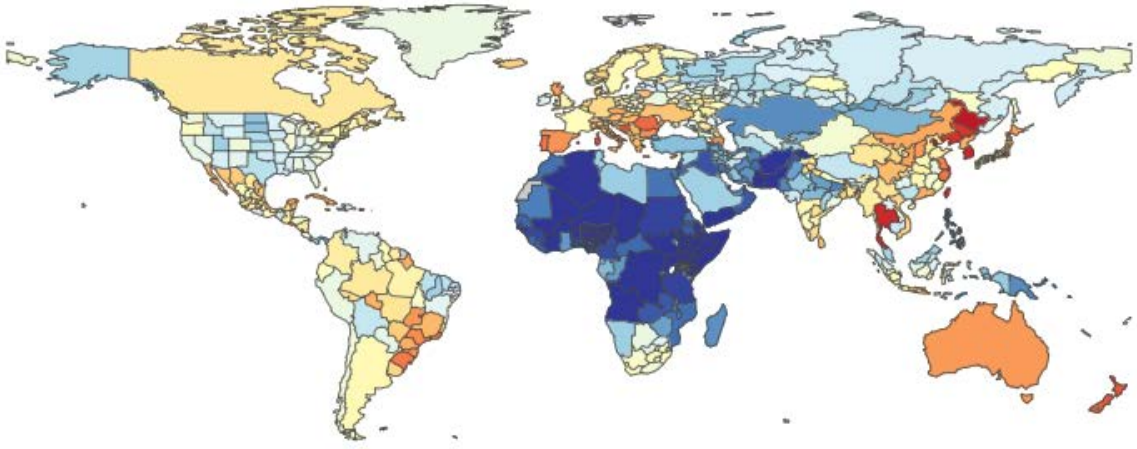


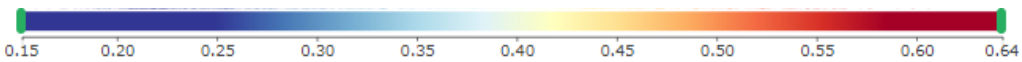

**Figure 4. Geographic distribution of firstborn prevalence in 1990 and 2019, according to the age-specific model.**



A. 1990



B. 2019



Quantitative metrics of model fit and bias for each of the two approaches are presented in Table 2. All three metrics were improved in the age-specific model, most especially data coverage within the 95% UI, which increased from 28.7% in the all-age model to 79.0% in the age-specific model.

**Table 2. Metrics of model performance for the all age and age-specific models**

| Model Version | RMSE | Mean Relative Error | Data Coverage |
|---|---|---|---|
| All-age model | 0.0244 | -3.55% | 28.7% |
| Age-specific model | 0.0197 | -2.85% | 79.0% |

Figure 5 shows a direct comparison of the prevalence results from the two models. While the global prevalence was similar, certain super-regions were significantly different. Firstborn prevalence in North Africa and the Middle East was lower in the all age model in 1990, Central Europe, Eastern Europe and Central Asia were lower in the all age model in 2019, and Sub-Saharan Africa and South Asia were lower in the all age model in both years.

**Figure 5. Firstborn prevalence in 2019 across all maternal ages, results from the all-age model versus the age-specific model.**

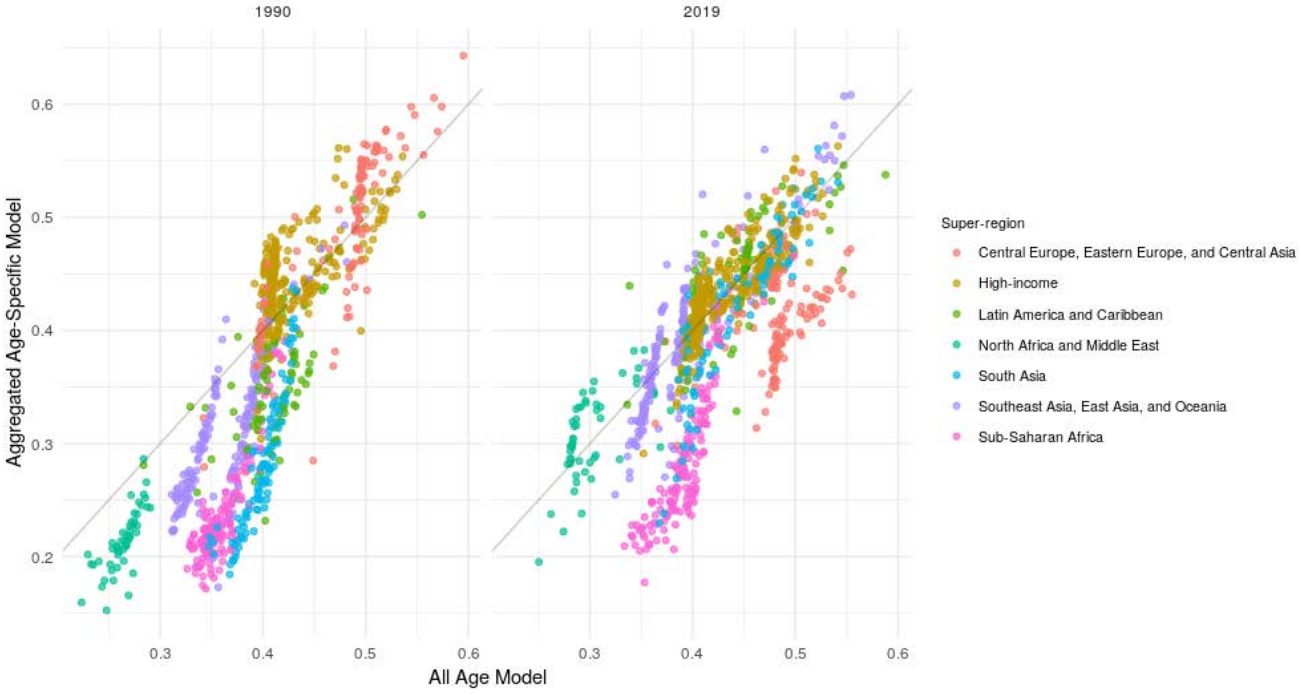
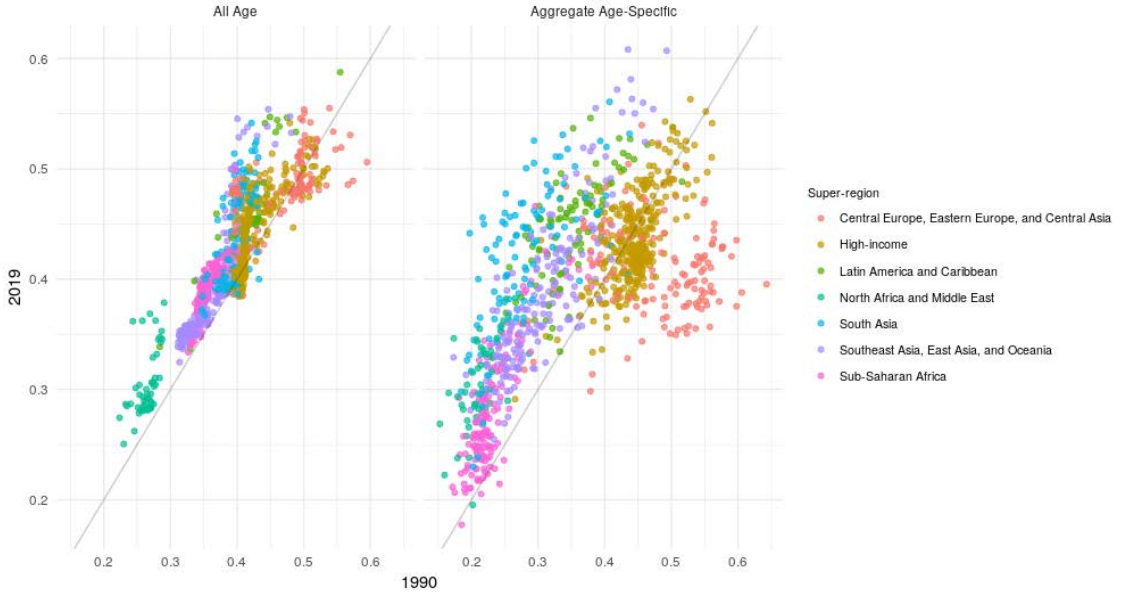


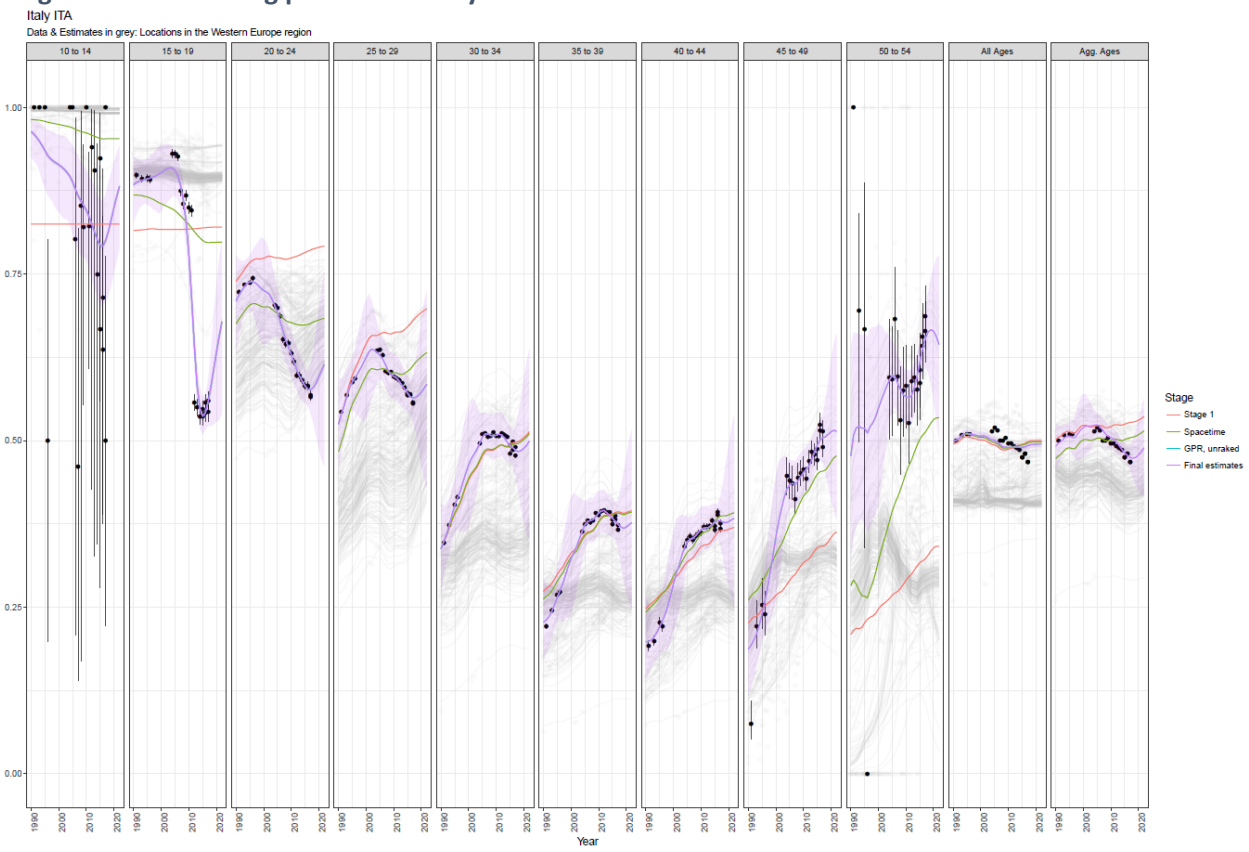
Figure 6 shows the difference between the two models in terms of time trend between 1990 and 2019. In the all age model on the left, firstborn prevalence increased in most locations, with the exception of some high-income countries and some countries in Eastern Europe. In the age-specific model on the right, the geographic pattern of which locations increased and which decreased remained largely the same, but the absolute change over time was larger than in the all age model.

**Figure 6. Change in firstborn prevalence from 1990 to 2019 in the all age model and the aggregate age-specific model.**




In addition to quantitative metrics of comparison, visualizations were generated for every location to qualitatively assess model performance. The age-specific model followed the data more closely in many locations. One example shown here is Italy, where Figure 7 shows the age-specific model (last panel) has objectively better fit to temporal changes in firstborn prevalence than are captured by the all age model (second to last panel).

**Figure 7. Firstborn prevalence by maternal age over time, including the input data and the three stages of the modeling process for Italy.**




*Prevalence Estimates*

The final model used the settings of the age-specific model described above with an expanded dataset that included any data extracted by maternal age group, regardless of whether the full maternal age range was covered. For comparability, the age-specific results were also aggregated into an all age summary. In the final model, global firstborn prevalence in 2019 was estimated as 38.9% (36.9 – 40.9), an increase of 8.72% (1.50 – 16.8) since 1990. This was a slightly higher prevalence than in either of the first two models, with an amount of increase since 1990 in between the first two models. The geographic distribution of firstborn prevalence in 1990 and 2019 is shown in Figure 8 (both sets of results are shown aggregated across age).

**Figure 8. Geographic distribution of firstborn prevalence in 1990 and 2019, according to the age-specific model run on all data.**

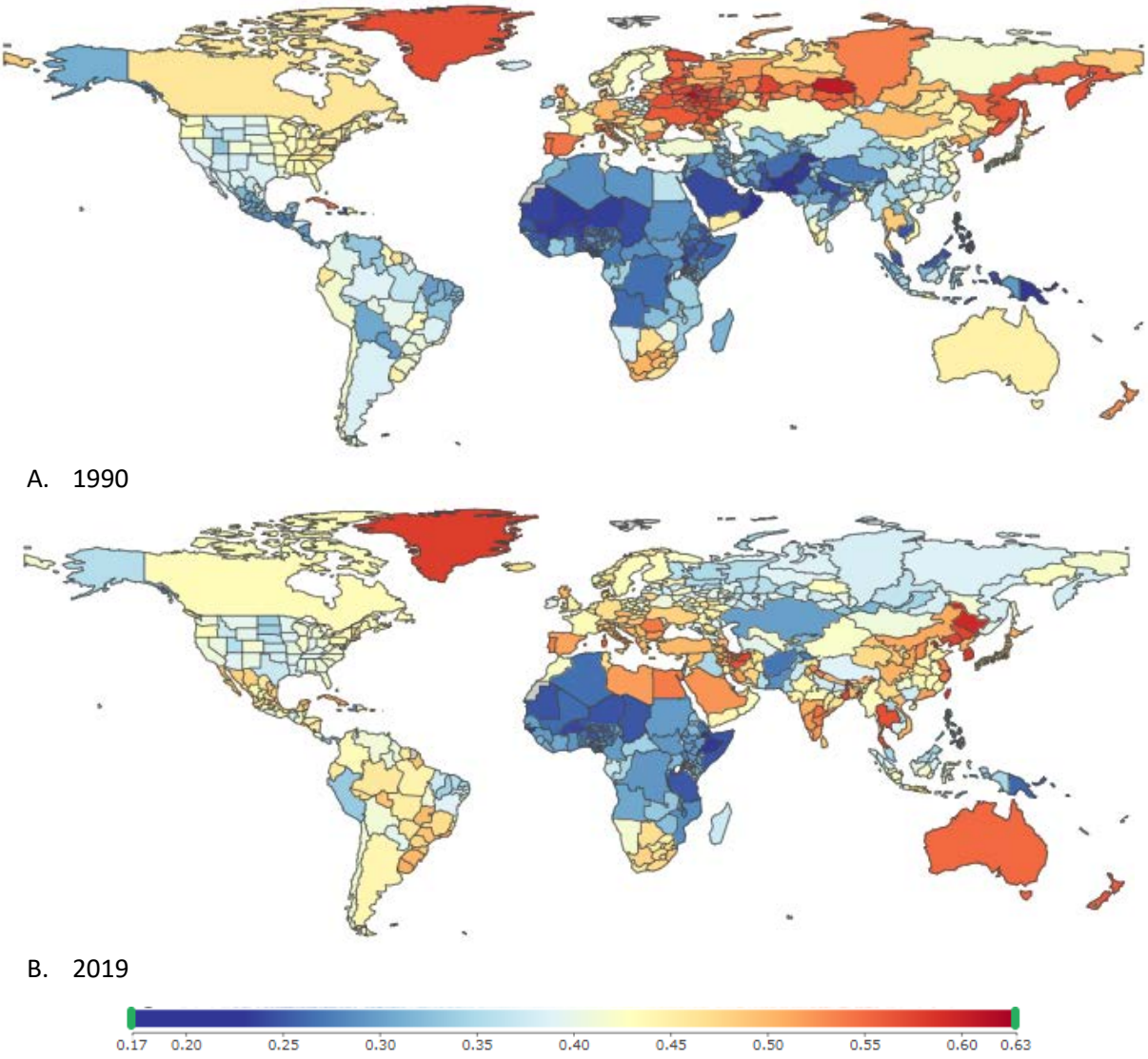
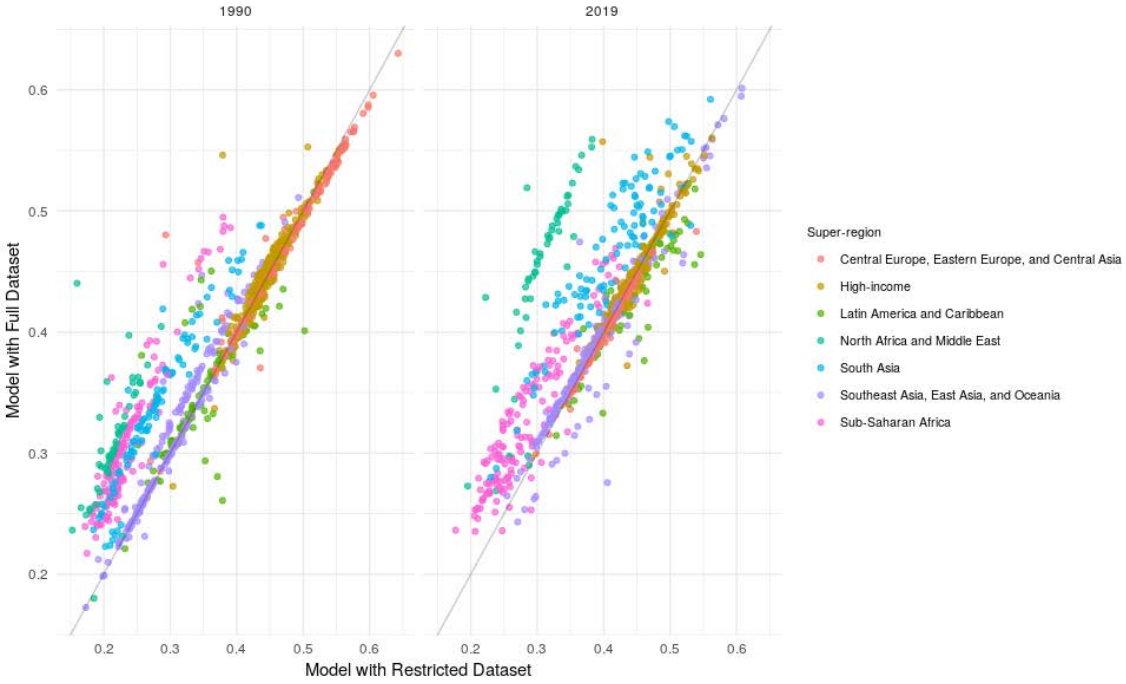


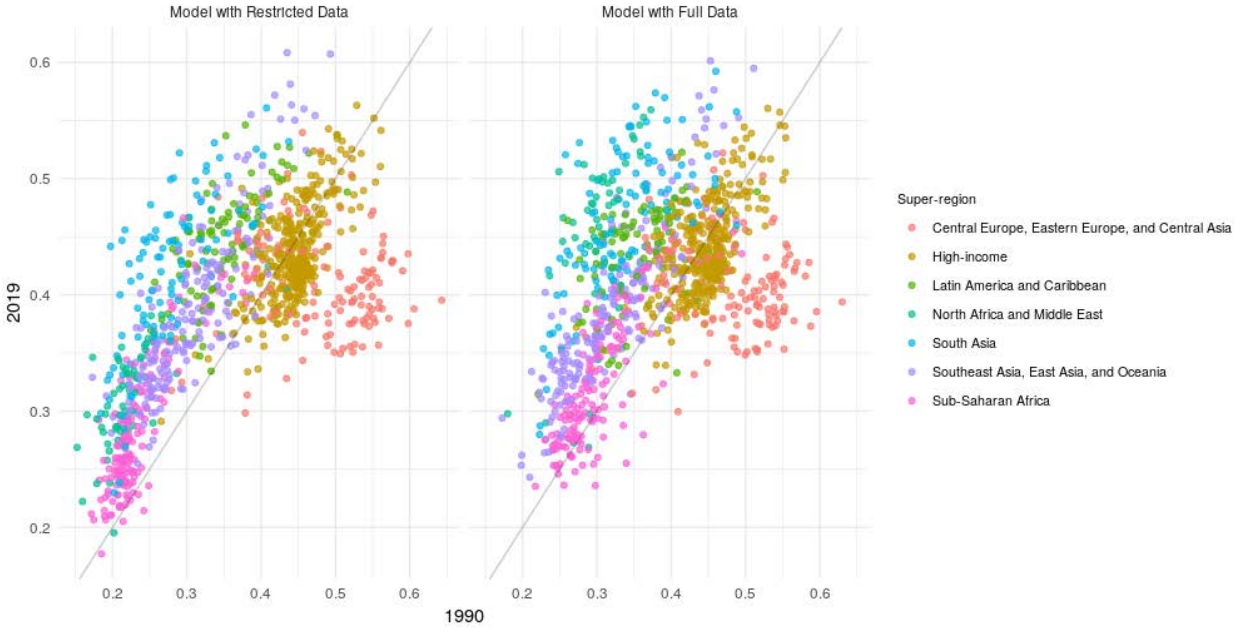

Figure 9 directly compares the prevalence results from the two models. Locations with complete coverage of the maternal age range changed very little (as in the High-income and Central Europe, Eastern Europe and Central Asia super-regions). Higher results were observed for locations where data with some missing age groups was now included (North Africa and the Middle East, Sub-Saharan Africa, and South Asia).

**Figure 9. Comparison of firstborn prevalence in 1990 and 2019 between the age-specific model run on the restricted and the full datasets.**



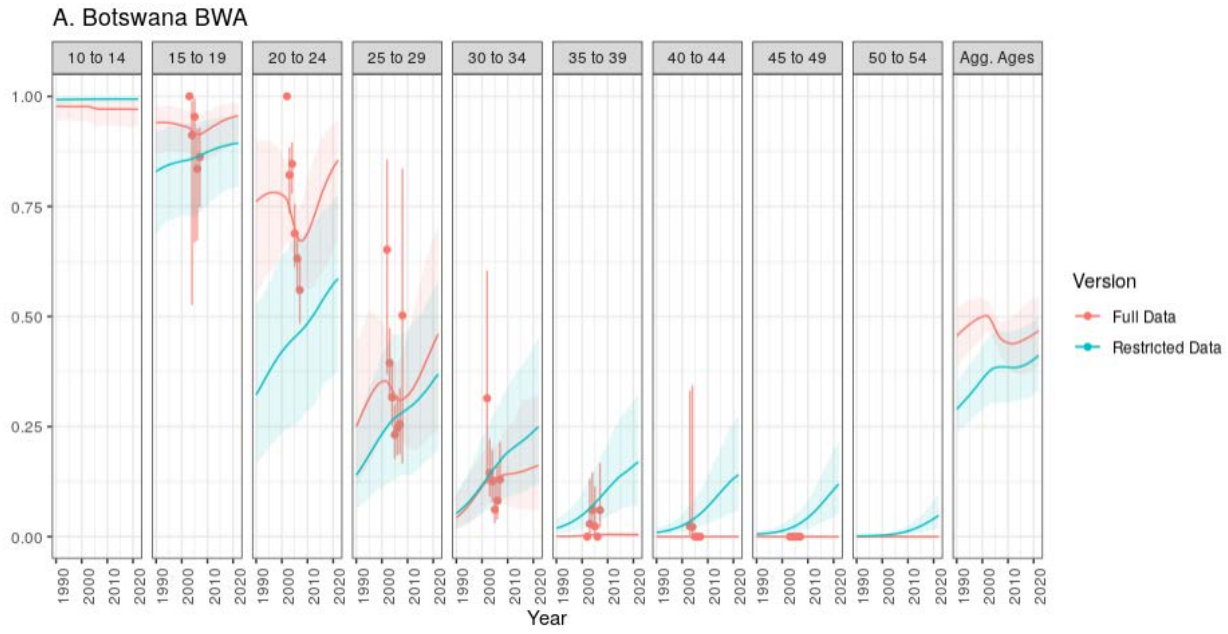

For temporal trends, Figure 10 shows the geographic distribution of change over time in firstborn prevalence, which was largely unchanged by the inclusion of additional data. The level of the firstborn prevalence was higher in some locations, such as in Sub-Saharan Africa, but the amount of change over time was similar.

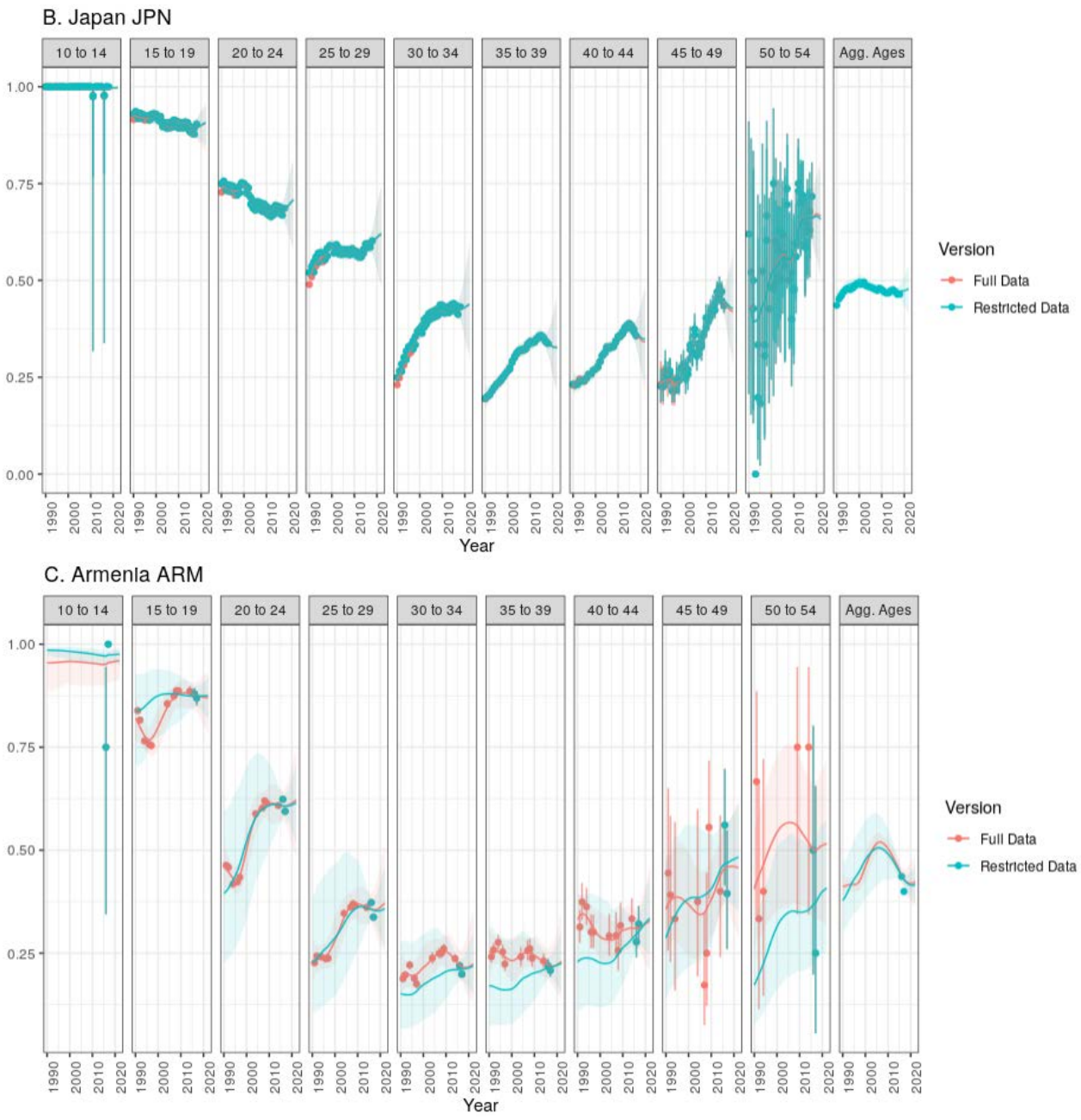
**Figure 10. Comparison of change in firstborn prevalence over time in the age-specific model run on the restricted and full datasets.**



Locations that diverged most from the model run on the restricted data were locations that went from no data in the restricted dataset to several years of data in the expanded dataset. The largest changes were observed in Yemen, Egypt, Botswana and Mongolia, all of which had no data in the restricted dataset (Botswana shown in detail in Figure 11A). In the two super-regions of North Africa and the Middle East and Sub-Saharan Africa, where many other countries still did not have data, these country-level changes also increased the prevalence estimates throughout the super-region. In locations with very complete data, such as Japan, the estimates remained highly similar (Figure 11B). In some locations where new years of data were included in the full dataset, like Armenia, the aggregate estimate remained similar, but estimates within maternal age groups changed (Figure 11C). Figure 11 shows examples of these three types of changes by showing the firstborn prevalence over time for each maternal age group and the aggregate, with the full data model in red and the restricted data model in blue, for Botswana, Japan, and Armenia.

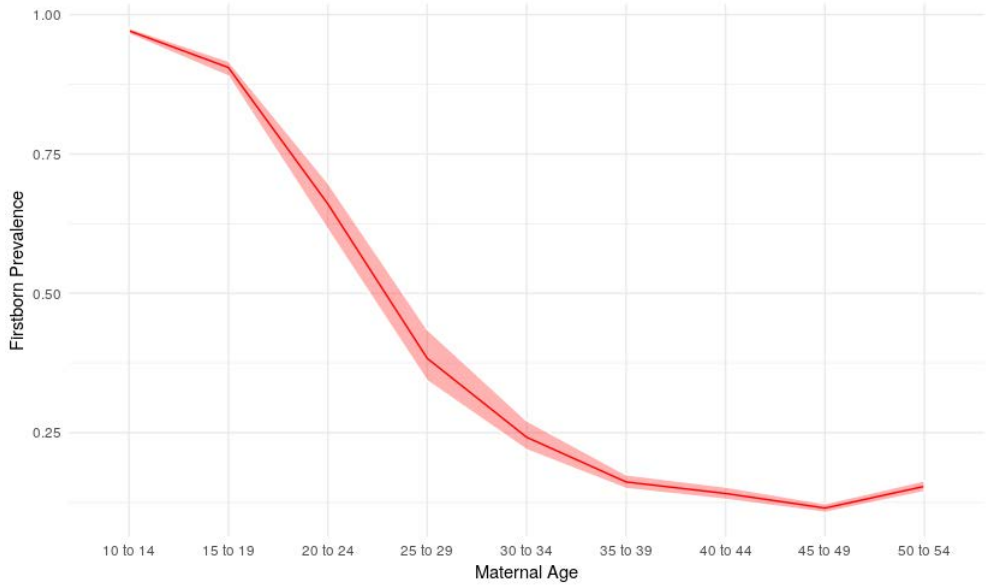
**Figure 11. Comparison of the impact of the full dataset on firstborn prevalence trends in three countries with varying degrees of impact, Botswana (high impact), Japan (low impact), and Armenia (intermediate impact). Data points in red were only present in the full dataset model; data points in blue were present in both models.**






Modeling by maternal age group also allowed examination of trends in firstborn prevalence by maternal age. Globally in 2019, firstborn prevalence was highest in the youngest age group of 10-14 years old and lowest in the second oldest age group of 45-49 years old (Figure 12).

**Figure 12. Firstborn prevalence by maternal age, globally in 2019.**




## Discussion

This analysis introduces a new modeling approach to estimate firstborn prevalence annually from 1990 to 2022, at the global, regional, national, and selected subnational levels, with a focus on accounting for maternal age composition and age-specific temporal fertility trends. When the model using CCFR was run on the same dataset as the model using traditional TFR, the global, all age firstborn prevalence was similar (35.1% versus 37.9% in 2019). However, the age-specific model is a clear improvement for three main reasons. First, in terms of sample-predictive performance, the age-specific model performed better than the all age model in the three metrics tested: RMSE, MRE, and data coverage. Based on MRE, both models had a negative bias, with less bias in the age-specific model. Better in-sample performance was further supported by qualitative examination of fit to data in individual locations. In many locations, the aggregated age-specific estimates have a better fit to the data and more accurately capture the time trends expressed in the data (Figure 7). A possible explanation for this improved fit over time is that by incorporating the fertility rates experienced by a maternal cohort over time instead of the fertility rates in a single year, the age-specific analysis better captures the impact of past fertility rates on the firstborn prevalence in the current year.

Second, the framework of modeling by maternal age better supports use of firstborn prevalence data collected over different maternal age ranges. This analysis highlighted that many

lower income countries collect birth order data from mothers 15-49 years old (i.e. Botswana in Figure 11), while many higher income countries collect birth order data from mothers 10-54 year old (i.e. Japan and Armenia in Figure 11). When modeling across all ages, the options are to define the age range for the analysis and exclude data with different age coverage, or to make the assumption that the different age representation does not matter and treat all data equally in the analysis. The first option limits the amount of data used to inform the model, and the second relies on an assumption that could easily be violated. In contrast, the age-specific framework supports inclusion of data that covers different maternal age ranges, without assuming every maternal age range is interchangeable.

Third, modeling by maternal age is more informative and more actionable. Firstborn prevalence changes significantly across maternal age groups (Figure 12), meaning outcomes associated with firstborn prevalence are also likely to vary significantly with maternal age. The age-specific estimates of firstborn prevalence can be used as more informative predictive covariates in analyses of other health outcomes or demographic trends, such as in the motivating problem of modeling prevalence of neonatal hemolytic disease due to Rh disease. The framework for modeling firstborn prevalence by maternal age is also more actionable than modeling across all ages because it allows for greater decomposition of what drives changes in firstborn prevalence, whether it is age composition, or temporal trends in fertility, or both. The greater specificity of the age-specific model supports more focused hypothesis generation. The age-specific modeling framework can also be used to model different fertility scenarios and compare the outcomes.

There are several important limitations to this work. While modeling by age has the ability to address and account for important challenges in modeling firstborn prevalence and birth order overall, it also presents and reveals some additional challenges. First, in order to accurately compare the all-age and age-specific models by using equivalent data, both models had to be run on only data that covered the full maternal age range of 10-54 years old. This restriction meant most of the data from lower income countries was excluded, and that modeling prevalence in these countries relied heavily on the fertility covariate and borrowing strength from other locations. The comparison between the two approaches remains valid since error due to the limited data should be present in both models, but an ideal situation would be to be able to compare equivalent data from a more geographically representative dataset. Second, additional data standardization

and imputation methods are needed to fully utilize data that is collected in age bins wider than five years. Age-splitting methods would need to be developed to distribute data that currently exists in age bins wider than five years into the standard five year bins used in this analysis. These efforts will be required to faithfully extend this approach to generate a full set of national, regional, and global estimates. Third, this analysis focused on the in-sample predictive performance of the two modeling approaches and did not test a wide array of candidate covariates based on out-of-sample predictive validity, approaches that could certainly be considered to increase robustness of the estimates. Fourth, the fertility covariates used in these analyses were estimates themselves and have their own limitations in terms of data availability. In general, locations with high-quality data on firstborn prevalence also have high-quality data on fertility, and locations with limited data on firstborn prevalence also have limited data on fertility. This suggests estimates of firstborn prevalence that rely heavily on the fertility covariate due to limited firstborn prevalence data may still be inaccurate due to limited underlying fertility data. Lastly, when considering the application of using firstborn prevalence to model Rh disease, one limitation is that it is possible for a first stillbirth to trigger the maternal immune response, and stillbirths are not included in this analysis. However, firstborn prevalence using live births is a very good proxy for who is at risk of Rh disease, and data on stillbirths is much more limited, especially in surveys.[17]

The estimates produced in this analysis, as well as the modeling framework itself, have many applications in the fields of health metrics and demography. To further increase the value of these estimates, future work should include testing out-of-sample predictive validity and optimizing model parameters around that goal, as well as developing methods for handling wide maternal age bins to maximize the quantity and geographic and temporal coverage of birth order data. Interesting future work would be to attempt to model birth order as a numeric variable rather than a binary of firstborn or later born. The small numbers in the data once it is stratified by birth order in addition to location, year, and maternal cohort present a major challenge. Exploration of the impact of birth spacing and maternal mortality on firstborn prevalence would be two additional interesting expansions on this work.

**References**


1 Mishra SK, Ram B, Singh A, Yadav A. Birth order, stage of infancy and infant mortality in India. *J Biosoc Sci* 2018; **50**: 604–25.

2 Gavrielov-Yusim N, Battat E, Neumann L, Friger M, Balicer RD. Birth order and private voluntary immunization--a study of 110,902 children. *Vaccine* 2012; **30**: 442–7.

3 Behrman JR, Taubman P. Birth order, schooling, and earnings. *J Labor Econ* 1986; **4**: 121–50.

4 Easey KE, Mars B, Pearson R, Heron J, Gunnell D. Association of birth order with adolescent mental health and suicide attempts: a population-based longitudinal study. *Eur Child Adolesc Psychiatry* 2019; **28**: 1079–86.

5 Moise KJ. Management of rhesus alloimmunization in pregnancy. *Obstetrics & Gynecology* 2002; **100**: 600–11.

6 James SL, Abate D, Abate KH, *et al.* Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet* 2018; **392**: 1789–858.

7 Corr P, Kippen R. The Case for Parity and Birth-Order Statistics. *Australian & New Zealand Journal of Statistics* 2006; **48**: 171–200.

8 GBD 2019 Demographics Collaborators. Global, regional, and national age-sex-specific fertility, mortality, and population estimates, 1950–2019: a comprehensive demographic analysis for the Global Burden of Disease Study 2019. *The Lancet* Forthcoming.

9 United Nations Population Division | Department of Economic and Social Affairs. https://www.un.org/en/development/desa/population/publications/dataset/fertility/age-fertility.asp (accessed Aug 12, 2020).

10 United Nations Statistics Division - Demographic and Social Statistics. https://unstats.un.org/unsd/demographic/products/dyb/default.htm (accessed Aug 20, 2020).

11 Human Fertility Database. https://www.humanfertility.org/cgi-bin/main.php (accessed Aug 20, 2020).

12 The DHS Program - Quality information to plan, monitor and improve population, health, and nutrition programs. https://dhsprogram.com/ (accessed Aug 20, 2020).

13 Home - UNICEF MICS. http://mics.unicef.org/ (accessed Aug 20, 2020).

14 Reproductive Health Surveys | Publications | Reproductive Health | CDC. 2019; published online Jan 16. https://www.cdc.gov/reproductivehealth/global/tools/surveys.htm (accessed Aug 20, 2020).

15 Vos T, Barber RM, Bell B, *et al.* Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *The Lancet* 2015; **386**: 743–800.

16 Zheng P, Aravkin AY, Barber R, Sorensen RJD, Murray CJL. Trimmed Constrained Mixed Effects Models: Formulations and Algorithms. *arXiv:190910700 [math, stat]* 2019; published online Sept 23. http://arxiv.org/abs/1909.10700 (accessed Aug 21, 2020).

17 Lawn JE, Blencowe H, Waiswa P, *et al.* Stillbirths: rates, risk factors, and acceleration towards 2030. *The Lancet* 2016; **387**: 587–603.