

UC Riverside

UC Riverside Previously Published Works

Title

Estimating the stochastic uncertainty in sample-based estimates of infant mortality in Ghana

Permalink

<https://escholarship.org/uc/item/2k72x1w3>

Journal

Journal of Economic and Social Measurement, 44(2-3)

ISSN

0747-9662

Authors

Kposowa, Augustine

Baker, Jack

Swanson, David A

Publication Date

2020-01-03

DOI

10.3233/jem-190462

Peer reviewed

Estimating the stochastic uncertainty in sample-based estimates of infant mortality in Ghana

Augustine Kposowa^a, Jack Baker^b and David A. Swanson^{a,c,*}

^a*Department of Sociology, University of California Riverside, Riverside, CA 92521, USA*

^b*Transamerica Life, Cedar Rapids, IA 52404, USA*

^c*Center for Studies in Demography and Ecology, University of Washington, Seattle, WA 98195, USA*

The Infant Mortality Rate (IMR) is an important population health statistic often used as one of the indicators of the health status of a nation. In many countries lacking adequate vital registration systems, sample methods are used to estimate IMRs. However, evaluations of this approach are rare and the literature contains no assessments of the stochastic uncertainty underlying these estimated IMRs. Stochastic uncertainty reflects the fact that even where the underlying IMR is constant in a small population over time, there is a likelihood of yearly fluctuation in its empirical observations even if it is measured from a complete count of the events of interest. In this study a method is presented that can be used to assess this stochastic uncertainty. We use the country of Ghana as a case study for this purpose. The method, a beta-binomial model, is described, tested for validity, and illustrated using 2014 sample-based estimates of IMR for 13 sample regions in Ghana. As such, the approach we described regarding the revision of sample-based IMR estimates is aimed at taking into account of the stochastic uncertainty while preserving the information concerning the uncertainty due to sampling. In applying the method to Ghana, we find that the sample-based IMR estimates perform well in accounting for stochastic uncertainty and could be applied elsewhere.

Keywords: Representational data, estimate, Beta Binomial model, validity

1. Introduction

Infant mortality level is one of the most important indicators of the state of health and quality of life in any nation [1–5].¹ Recent data show that worldwide, the infant

*Corresponding author: David A. Swanson, Department of Sociology, University of California Riverside, Riverside, CA 92521, USA. Tel.: +951 827 6398; E-mail: dswanson@ucr.edu.

¹Murray [35] has argued that the infant mortality rate is flawed when it is used as an index of overall mortality (i.e., the mortality regime affecting a given population) and that “Disability Adjusted life Expectancy” should be used in its place. However, it has been pointed out by Reidpath and Allotey [36] that the infant mortality rate and the measure proposed by Murray are so highly correlated that it merely goes to reinforce the intuition that the causes of infant mortality are strongly related to those structural factors like economic development, general living conditions, social well-being, and environmental factors, and, and such, the infant mortality rate remains a useful and comparatively inexpensive indicator of population health. Guillot et al. [37] also note that infant mortality is responsive to changes in annual mortality conditions because it involves a short lag between the timing of mortality exposures and the timing of corresponding births.

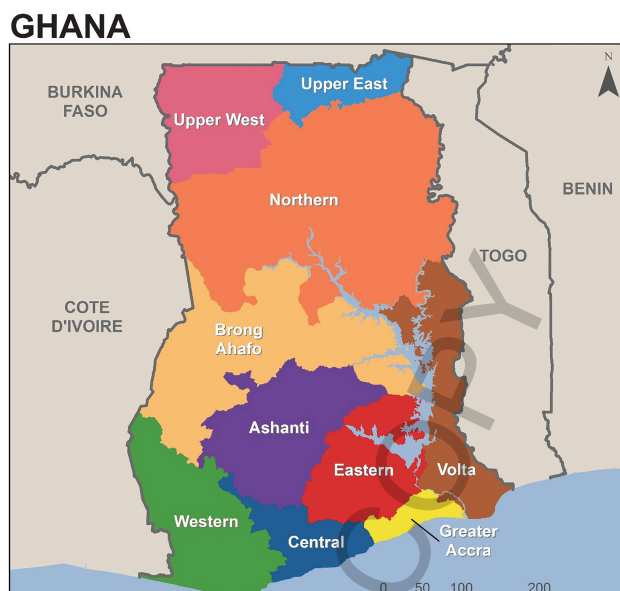


Fig. 1. Map of Ghana and its ten administrative regions. Source: Ghana statistical service and ICF international, 2015: xxii.

mortality rate (IMR) has declined since 1990. It went from 64.8 deaths per 1,000 live births in 1990 to 30.5 deaths per 1,000 live births in 2016, representing a decrease of over 52% [6]. Despite the decline, there remain substantial IMR variations by world regions, with countries in sub-Saharan Africa experiencing the highest rates. Within sub-Saharan Africa, Ghana has had one of the lower rates, though by global standards, its infant deaths remain relatively high. It had an estimated rate of 53.4 deaths per 1,000 live births in 2008, which declined to 41.2 deaths per 1,000 live births in 2016 [7]. There is evidence to suggest that the relatively high infant deaths in Ghana are in part attributable to the country's high neonatal mortality rate [8].

As shown in Fig. 1, there are 10 administrative regions in Ghana: Western, Central, Greater Accra, Volta, Eastern, Ashanti, Brong Ahafo, Northern, Upper East, and Upper West [9]. Current national borders were established in the 1900s when Ghana was known to Europeans as the British Gold Coast. It became independent in 1957 and became a republic in the British Commonwealth in 1960 [9]. Ghana's population was estimated at 27 million in 2014 [9]. The Ashanti, Eastern, and Greater Accra regions together constitute about 50 percent of the country's population. Upper East is the least populated region, accounting for two percent of the total population of Ghana [9]. The 2014 Demographic and Health Survey of Ghana showed evidence that there are substantial regional variations in infant mortality [9]. Nationally, it provided an estimate of 41.2 deaths per 1,000 live births [9]. Greater Accra, the most developed region of the country and the nation's most populous area had a rate of

36.7, the lowest, followed by Volta (41.9). Moving further from Accra on the Atlantic coast, there seems to be a south-north gradient in mortality, which is likely reflective of rurality, ethnic differences, and especially socio-economic conditions [9, p. 103–104]. Ashanti experienced 63.5 deaths per 1,000 live births; the Upper East had 46.1, and the highest rate was reported in the Upper West (64.1 deaths).

Data availability and quality remain a major problem for any demographic analysis of infant mortality in Ghana [8–14]. Although Ghana has made numerous positive strides, in education, health, and infrastructure, the country lacks a fully functioning vital registration system of births and deaths. Estimates of births and deaths are based on surveys and hospital records [9–14].

Sample based estimates of infant deaths in Ghana for 2014 were obtained by asking women age 15–49 about their full birth history, including number of children that had died the previous year, and for each death, age at death [9]. Such sample based estimates are subject to error and sampling variability because survey participants are likely to underreport infant deaths particularly if such deaths had occurred some years ago. In African tradition, women do not want to remember their dead children and as such, many dead children are missed out of the estimate. This has implications for the numerator of the equation often used for determining mortality estimate. Furthermore, given high mortality at all ages, and high maternal mortality rates in Ghana [9,11–15], figures on infant mortality will by design be underestimated as women that died before the survey could not be asked about their own birth histories. Another source of potential bias is that women aged below 15 years and those above 49 years are not included in the sample, both of which tend to have higher infant mortality rates than those aged 15–49 [9, p. 104–105]. As the Ghana Statistical Service et al. [9, p. 100] acknowledges, inherent uncertainty exists in the use of surveys for the estimation of infant mortality due to possible biases arising from incomplete and possibly unrepresentative data.

In view of incomplete vital registration systems in Ghana, researchers studying infant mortality only rely on estimates that are sample-based. The yet unanswered question is: how reliable are the estimates of infant mortality obtained from sample based data, and to what extent could they be extrapolated to the entire population? Responding to these questions require some robust statistical techniques that are capable of providing valid estimates. The method we describe in this paper is therefore tested for validity and illustrated in a case study using 2014 sample-based estimates of IMR for the nation as a whole, its ten administrative regions an urban sample area, and a rural sample area [9]. The assumption underlying the use of this method is that there are two sources of variation in sample-based estimates of a variable, namely, sample size and the variation of the variable in question in the population from which the sample is taken. As such, the approach we describe in more detail in regard to revising sample-based IMR estimates is aimed at taking into account the stochastic uncertainty while preserving the information concerning the uncertainty due to sampling. In the process of accounting for stochastic uncertainty, our method also may

serve to ameliorate the sources of potential error that affect sample-based IMR estimates. In general, our method represents a “representative context” approach, based on the idea that it is a statistical estimator. To this end, a publication by Link and Hahn [16] was used as a guide in generating the approach described, tested, and applied.

Before turning to a description of our method and the data we employed, it is appropriate to provide an overview of what our approach aims to do. Stochastic uncertainty reflects the fact that even where the underlying IMR is constant in a small population over time, empirical observations of it can fluctuate from year to year even if it is measured from a complete count of the events of interest [17]. As such, the approach described regarding to revising sample-based IMR estimates is aimed at taking into account stochastic uncertainty while preserving the information concerning the uncertainty due to sampling. Note that we employ the definition that: (1) stochastic uncertainty is the manifestation of a process representing numerical values of some system randomly changing over time [18]; and (2) sampling uncertainty is an estimate attached to a test result that characterizes the range of values within which the true value is asserted to lie [19].

2. Methods

2.1. Infant mortality rate as a beta binomial process

The infant mortality rate measures the proportion of births that result in deaths during the first year of life. As such, it measures the relationship between events (deaths) and trials (births) where

$$d_{i,t} = D_{i,t}/B_{i,t}$$

where

$$i = \text{area}(i = 1 \text{ to } n)$$

$$t = \text{time}$$

$$D = \text{infant deaths}$$

$$B = \text{births}$$

The distribution of infant deaths in a given area i at a given time t is (approximately) binomial, with parameter d , and is typically described as a beta binomial random process with a probability mass function defined by two parameters: α and β . The first parameter, α , can be interpreted as the count of the event of interest, which in our case is the number of infant deaths, the number of births in which the infant dies before achieving the first year of life. The second parameter, β , can be

interpreted as the count of “non-events,” which in our case is the number of children born who survive to reach one year of age. Note that “rate” = $\alpha/(\alpha + \beta)$, which in our case is equivalent to “infant mortality rate” = infant deaths/(infant deaths + survivors to age 1), which reduces to infant deaths/births. Thus, parameter α is the numerator in the expression defining a rate, and when added together, the parameters α and β represent the denominator. Together, we may re-express the IMR rate as the compound distribution of α and β captured in the beta-binomial probability model:

$$\text{IMR} = \alpha/(\alpha + \beta) = \text{infant deaths}/(\text{infant deaths} + \text{infant survivors})$$

Because IMR may be conceptualized directly using the beta-binomial process, IMRs may be thought of as stochastic processes that occur within each region while also contributing to higher-level meta-populations within which they are nested [20,21]. As a general description, the method we employed is a distribution of IMRs estimated from a “representative context” from which the two parameters of a “Beta model” fitted to a binomial distribution made up IMRs (where the IMR is divided by 1,000) are used to develop an IMR estimate for populations of interest. Even where it is the case that the population is so small that neither infant deaths nor births are reported, the two parameters can be used to develop an estimate of the underlying IMR. However, in such a case, the underlying estimated IMR is an “average” based on the parameters generated by fitting a Beta Model to a distribution of IMRs selected as representative of the population(s) of interest.

2.2. An indirect estimator of IMR using averaging of samples from a beta-binomial stochastic process

A potential number of strategies exist for dealing with small sample size dynamics (or confidentiality suppression) in making estimates of infant mortality rates. First, one might simply use the IMR for the entire set (e.g., the nation as a whole) of hierarchical geographies (the provinces of a given nation) in place of highly-uncertain estimates of IMR for subsets of the whole, a strategy employed in the field of small area population estimates [22]. This would stabilize estimates for IMR for selected subsets, but at the expense of potentially masking heterogeneity in IMRs across geographic units. For purposes of capturing spatial patterns in IMR, a main priority in smaller-level analyses, this solution is less acceptable. A second alternative might be to make adjustments based on judgment. While this may improve estimates overall, especially when judgments are made by applied demographers with significant experience, this approach is subject to the criticism that non-standard methods are applied across different geographies and/or population groupings. An ideal approach would be to utilize a principled method for adjusting local estimates of IMR. Simple model averaging, based on the beta-binomial model represents a viable approach for achieving this goal.

We have established that the IMR constitutes a beta-binomial probability process. We may think of two estimates of this process as constituting samples of the mean

and variance of the underlying process. Therefore, we may consider these as samples obtained from the same underlying mortality process and in averaging them, we might anticipate arriving upon a superior estimate of the mean proportion [20,21]. As such, the averages of two estimates based on the model may also be averaged as:

$$\text{IMR}_{\text{averaged}} = (\alpha_1 + \alpha_2) / ((\alpha_1 + \beta_1) + (\alpha_2 + \beta_2))$$

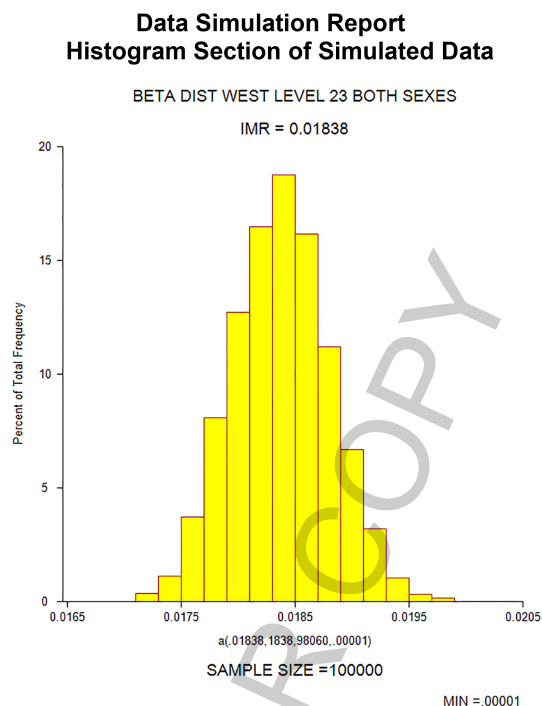
where the subscripts (1, 2) now represent estimates of death and survivorship counts for two groups. This method can, of course, be extended to k groups as desired. Such model averaging yields an estimate where a larger-scale and representationally-appropriate model IMR is leveraged to make smaller-scale estimates more precise in a manner similar to that observed in the literature on indirect estimation in demography [23,24]. Recent attempts to extend indirect estimation based on stochastic process theory have been introduced [25] and here, we leverage this idea further in developing indirect estimates of IMR. We note that this approach to estimating IMR appears to be new in that no description of it is found in the literature [24,26,27].

Before turning to a discussion of the data, it is appropriate to discuss in some detail the averaging process just described. Because an IMR is typically expressed per 1,000 births, it can be turned into a binomial variable by dividing it by 1,000 (or more generally if IMR is expressed as infant deaths per k births, it would be divided by k). In this form, IMR is strictly bound in that it cannot be less than zero nor greater than 1 ($0 \leq \text{IMR} \leq 1$). Because the denominator of the IMR measure contains those at risk of dying, it is a probability. In practice, it is substantially less than one.

2.3. Validity test

Given that our proposed method is producing a revised IMR that is likely to be close to the underlying IMR for a small population and therefore reflective of its intrinsic mortality regime, one would expect the method to do this where one could observe the intrinsic mortality regime. Model stable populations afford this opportunity because they have known intrinsic mortality regimes, the model life tables associated with a given set of model stable populations. To examine how the method works in this environment, we employed the IMR associated with a model stable population found in Manual IV, Methods of Estimating Basic Demographic Measures from Incomplete Data [24]. For this purpose, we selected the infant mortality rate associated with West Level 23 for both sexes, which shows that of 100,000 births, 98,166 are expected to reach their first birthday. This yields an $\text{IMR} = 1 - 0.98166 = 0.0184$.

Using the IMR of 0.0184 and a seed population of 100,000, a random sample of 5,000 IMRs was generated using the beta-binomial model simulation provided by the NCSS statistical system (release 8). The sample is sufficiently large to allow the simulation program the opportunity to generate outliers, which it did. As can be seen in Fig. 2, the mean is 0.01838 with a standard deviation of 0.000423 and a coefficient of variation equal to 0.02305. The minimum IMR is 0.016849 and the maximum is 0.020147.



Descriptive Statistics of Simulated Data

| Statistic | Value | Statistic | Value |
|--------------------------|--------------|-----------------|------------|
| Mean | 0.01838248 | Minimum | 0.01684878 |
| Standard Deviation | 0.0004237251 | 1st Percentile | 0.01744547 |
| Skewness | 0.07195781 | 5th Percentile | 0.01769227 |
| Kurtosis | 2.934381 | 10th Percentile | 0.0178332 |
| Coefficient of Variation | 0.02305049 | 25th Percentile | 0.01808544 |
| Count | 5000 | Median | 0.01838394 |
| | | 75th Percentile | 0.01866444 |
| | | 90th Percentile | 0.0189272 |
| | | 95th Percentile | 0.01908542 |
| | | 99th Percentile | 0.01937772 |
| | | Maximum | 0.02014658 |

Fig. 2. NCSS report on the fit of the beta-binomial model to the IMRs of 58 counties in the validity test.

From the 5,000 randomly generated observations, we extracted two sets of data. For the first set, we extracted the initial 43 IMR randomly generated observations from the simulation. For the second, we rank-ordered the 5,000 observations: from high to low and then from low to high, and extracted the eight highest IMR and seven lowest IMRs, respectively from them. The idea is that the entire set represents a synthetic population with 58 observations while the second set of 43 sim-

ulated IMRs represents the subset of the synthetic population in which IMRs are reported, and the third set of 15 simulated IMRs represents a subset of “small populations” subject to a high level of stochastic uncertainty. These characteristics mimic the 2009–2011 IMRs reported for the 58 counties of state of California, where the results are not reported for 15 counties (due to their small populations). As such, the validity test was set up as if there were 43 units for which IMRs were reported and 15 for which they were not. However, all of the data used in the validity test were generated from the synthetic population that is based on Model Life Table, Level 23, as described in the text. The reporting structure as well as the actual data for California can be found through the Open Portal service provided by the California Health and Human Services Agency via a download of a CVS data set assembled by the California Department of Public Health. This data set can be accessed by going to: <https://data.chhs.ca.gov/dataset/infant-mortality-deaths-per-1000-live-births-lghc-indicator-01/resource/ae78da8f-1661-45f6-b2d0-1014857d16e3> and then clicking on the “download” tab, which downloads the file, “Infant Mortality, Deaths Per 1,000 Live Births (LGHC Indicator 01) (CSV)” in CVS form. Once downloaded, it can be saved as an excel file. The data in this file include the infant mortality rates (identified as “rate” in the file) and the infant deaths (identified as “numerator” in the file) and live births (identified as “denominator” in the file) used to calculate the IMRs for all counties and other administrative areas, including the state as a whole. The data represent the period 2009–11. A description of the methods, caveats, and so forth associated with this data set can be found on the ULR shown above.

The 43 observations are expected to be closer, on average, to the “underlying” IMR of 0.01838 or all 5,000 observations and have less variation, respectively, than that found in the 15 observations. For the set of 43 observations, the mean IMR is 0.01834 and the coefficient of variation is 0.02305. For the set of 15 observations, the mean IMR is 0.01855 and the coefficient of variation is 0.07692. Thus, the set of 43 observations has a mean and a coefficient of variation closer to the mean and coefficient of variation found in the full set of 5,000 observations than does the set of 15 observations.

A beta-binomial model was fit to the set of 43 observations and its parameters were used to revise the IMRs in the set of 15 observations. The expectation is that the revised IMRs will yield a mean IMR closer to that found for the full 5,000 set of simulated observations and that the variation among these revised means will decline, yielding a smaller coefficient of observation.

The results show that the model moved the initial IMR estimates for the 15 observations closer to the underlying IMR. As such, they are more reflective of the West Level 23 mortality regime that is intrinsic to them: the mean of the original IMRs for the 15 observations is 0.01855 while the mean for the revised IMRs is 0.01839, which is closer to the underlying IMR of 0.01838. In terms of variation, the coefficient of variation for the initial set of 15 IMRs is 0.07692, while that for the revised set is 0.00338. These results support the argument that the method described

in this paper is capable of moving IMRs subject to stochastic uncertainty closer to the underlying IMRs and their respective intrinsic mortality regimes.

In the validity test, different counties are simulated from a common beta distribution, and the result is that the two sets of counties, large and small, are normally distributed around the intrinsic mean IMR of the “population”. The simulation shows that the adjusted IMRs of the small counties move closer the intrinsic IMR, which indicates that the method works when both the small counties and large counties represent samples taken from the same underlying population. If the small counties represent a sample from a different population than the sample of large counties, then the adjustment may yield a “biased” estimate of the former’s intrinsic IMR. This shows the importance of having a reference set that conceptually represents a sample from the same underlying population as the small county sample. One way to visualize the unbiased and biased outcomes is to picture the case where the method yields: (1) an “unbiased” estimate, which is when the mean IMR of the large counties is between the intrinsic IMR of the counties and the small counties mean IMR; and (2) a “biased” estimate when the method does not move the mean IMR for the small counties closer to its intrinsic IMR, which occurs where the mean IMR of the small counties is between its intrinsic IMR and the mean IMR of the large counties.²

3. Results

Along with sample size and related information, Table 1 displays the 2014 survey-based IMRs for Ghana by the sampled regions, which include aggregations of the administrative districts and the districts themselves (9). As can be seen in Table 1, the national IMR is estimated to be 41.245, with a 95% confidence interval of 34.516–47.974. Among the ten regions, IMR is lowest for Greater Accra (36.680), with a 95% confidence interval of 22.772–50.588, while the highest is for Ashanti (63.479), which has a 95% confidence interval of 46.803–80.155.

Turning to the uncertainty in the sample-based IMR estimates, we discuss the standard errors, which provide information similar to that found using the coefficients of

²In the validity test, different counties are simulated from a common beta distribution, and the result is that the two sets of counties, large and small, are normally distributed around the intrinsic mean IMRs of the “population”. The simulation shows that the adjusted IMRs of the small counties move closer the intrinsic IMR, which indicates that the method works when both the small counties and large counties represent samples taken from the same underlying population. If the small counties represent a sample from a different population than the sample of large counties, then the adjustment may yield a “biased” estimate of the former’s intrinsic IMR. This shows the importance of having a reference set that conceptually represents a sample from the same underlying population as the small county sample. One way to visualize the unbiased and biased outcomes is to picture the case where the method yields: (1) an “unbiased” estimate, which is when the mean IMR of the large counties is between the intrinsic IMR of the counties and the small counties mean IMR; and (2) a “biased” estimate when the method does not move the mean IMR for the small counties closer to its intrinsic IMR because the mean IMR of the small counties is between its intrinsic IMR and the mean IMR of the large counties.

Table 1
2014 ghds IMR by region for ghana with sample characteristics*

| Sample area | Sample IMR | Standard error (sample IMR) | Lower 95% confidence limit of sample estimate of IMR (prior) | Upper 95% confidence limit of sample estimate of IMR (prior) |
|---------------|------------|-----------------------------|--------------------------------------------------------------|--------------------------------------------------------------|
| National | 41.245 | 3.433 | 34.516 | 47.974 |
| Urban | 49.057 | 4.493 | 40.251 | 57.863 |
| Rural | 46.108 | 2.852 | 40.518 | 51.698 |
| Western | 39.826 | 5.782 | 28.493 | 51.159 |
| Central | 48.001 | 6.207 | 35.835 | 60.167 |
| Greater accra | 36.680 | 7.096 | 22.772 | 50.588 |
| Volta | 41.876 | 8.391 | 25.430 | 58.322 |
| Eastern | 43.110 | 7.155 | 29.086 | 57.134 |
| Ashanti | 63.479 | 8.508 | 46.803 | 80.155 |
| Brong ahofo | 37.624 | 4.716 | 28.381 | 46.867 |
| Northern | 52.715 | 7.275 | 38.456 | 66.974 |
| Upper east | 46.248 | 8.743 | 29.112 | 63.384 |
| Upper west | 64.099 | 7.691 | 49.025 | 79.173 |

*Ghana Health and Demographic Survey 2014 (October, 2015), Ghana Statistical Service, Accra, Ghana. Ghana Health Service, Accra, Ghana. The DHS Program, ICF International, Rockville, MD USA. From The National Sample to the Upper West Sample, respectively, data (for women) are taken from tables, B2 through B.14. (<https://dhsprogram.com/pubs/pdf/FR307/FR307.pdf>).

variation. For example, the standard error for Volta (8.391) is over twice as high as the standard error for the country as a whole, at 3.433. The sample size (females) for the country as a whole (9,396) is just over twelve times the size of the sample for Volta (795). The relative population sizes are similar: In the 2010 census, the population for the country as whole (approximately 24.7 million) is approximately twelve times the size of that for Volta (approximately 2.1 million). Picking up on the discussion of sample uncertainty and stochastic uncertainty found at the end of the preceding section, we again note that there are two sources of variation in sample-based estimates of IMRs: (1) sample size and (2) the variation of infant deaths in the population from which the sample is taken. The standard errors reflect these two sources. Volta's estimated IMR also reflects a higher level of stochastic uncertainty because of the small size of its population. It is the latter that our approach is aimed at reducing in regard to revising sample-based IMR estimates, while at the same time preserving the information concerning the uncertainty due to sampling.

The Beta Binomial model procedure found within the "survival/reliability" module of the NCSS statistical analysis package (release 8) was used to obtain the two Beta Model parameters using the infant mortality rates found for 238 countries provided by the World Bank [15]. The major results of interest found in running this procedure with the data are found as Fig. 3. Note that there two different estimates of the α and β parameters, one accomplished by the method of moments and the other by Maximum Likelihood Estimation. The parameters of the latter are used here, namely: $\alpha = 1.241538$ and $\beta = 49.78341$. Again as stated earlier, we use the MLE approach because it comes with estimates of the standard error and confidence intervals.

Dataset C:\...\GHANA GDHS 2003\WORLD BANK 2015 imr by country.NCSS
 Time Variable imr

Parameter Estimation Section

| Parameter | Method of Moments Estimate | Maximum Likelihood Estimate | MLE Standard Error | MLE 95% Lower Conf. Limit | MLE 95% Upper Conf. Limit |
|----------------|----------------------------|-----------------------------|--------------------|---------------------------|---------------------------|
| Minimum (A) | 0 | 0 | | | |
| Maximum (B) | 1 | 1 | | | |
| α | 1.33442 | 1.241538 | 0.1020844 | 1.041457 | 1.44162 |
| β | 53.46343 | 49.78341 | 4.985078 | 40.01283 | 59.55398 |
| Log Likelihood | | -650.2952 | | | |
| Mean | 0.02435168 | 0.02433199 | | | |
| Median | 0.01883641 | 0.01843637 | | | |
| Mode | 0.006333965 | 0.004926846 | | | |
| Sigma | 0.0206349 | 0.02136161 | | | |

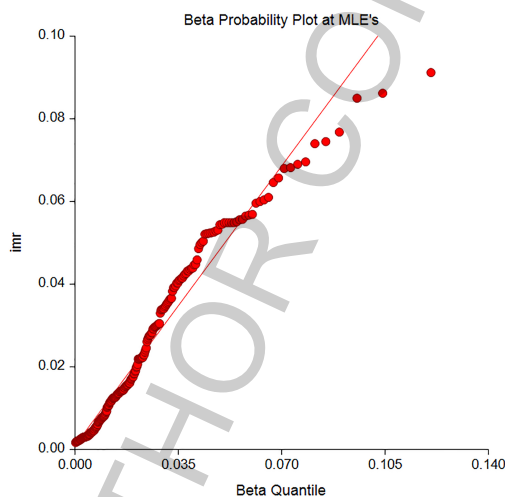


Fig. 3. NCSS report on the fit of the beta model to the IMRs of 238 countries.

Table 2 provides the revised 2014 IMRs estimated for each of the 13 Sampling Regions of Ghana using our procedure.

4. Discussion

Table 3 shows the change from the sample-based IMRs to the revised IMRs estimated by our method. In each case, the change resulted into a decline in the IMR. As summary statistics, we employ the Mean Algebraic Percent Error (MALPE) and the Mean Absolute Percent Error (MAPE). The former provides the average direction of change while the latter provides an assessment of the agreement between the two sets of estimates (22: 269–270). The MALPE for the all of the 13 sample regions

Table 2
Revised 2014 IMR estimates by region, Ghana

| Sample area | Sample IMR | Revised IMR estimates |
|---------------|------------|-----------------------|
| National | 41.245 | 40.858 |
| Urban | 49.057 | 48.723 |
| Rural | 46.108 | 45.842 |
| Western | 39.826 | 39.319 |
| Central | 48.001 | 46.991 |
| Greater accra | 36.680 | 36.309 |
| Volta | 41.876 | 41.097 |
| Eastern | 43.110 | 42.163 |
| Ashanti | 63.479 | 61.368 |
| Brong ahofo | 37.624 | 37.471 |
| Northern | 52.715 | 52.172 |
| Upper east | 46.248 | 45.029 |
| Upper west | 64.099 | 62.095 |

Table 3
Differences between the sample-based IMR estimates and the revised IMR estimates

| Sample area | Sample IMR | Revised IMR estimates | Difference (sample – revised) | Percent difference |
|---------------|------------|-----------------------|-------------------------------|--------------------|
| National | 41.245 | 40.858 | 0.387 | 0.95% |
| Urban | 49.057 | 48.723 | 0.334 | 0.68% |
| Rural | 46.108 | 45.842 | 0.266 | 0.58% |
| Western | 39.826 | 39.319 | 0.507 | 1.29% |
| Central | 48.001 | 46.991 | 1.010 | 2.15% |
| Greater accra | 36.680 | 36.309 | 0.371 | 1.02% |
| Volta | 41.876 | 41.097 | 0.779 | 1.90% |
| Eastern | 43.110 | 42.163 | 0.947 | 2.25% |
| Ashanti | 63.479 | 61.368 | 2.111 | 3.44% |
| Brong ahofo | 37.624 | 37.471 | 0.153 | 0.41% |
| Northern | 52.715 | 52.172 | 0.543 | 1.04% |
| Upper east | 46.248 | 45.029 | 1.219 | 2.71% |
| Upper west | 64.099 | 62.095 | 2.004 | 3.23% |

is 1.66%, indicating a slight downward direction in going from the sample-based IMRs to the revised estimated IMRs. The MAPE also is 1.66%, which indicates that the revised estimated IMRs are relatively close to the sample-based IMR estimates. Taken altogether, the results suggest that the stochastic uncertainty underlying the sample-based estimates is not huge and, as such, they perform well in accounting for it.

The estimates are subject to judgment, which largely occurs in the selection of the “representative context” used to construct the Beta Model and its two parameters. However, even still the entire process is transparent, which means that the results are not subject to arbitrary and capricious judgments that render them difficult to replication. This and the fact that estimates are valid and can be efficiently generated by the process described here suggests that they have the potential to assess the stochastic uncertainty underlying sample-based estimates of IMRs. Estimates of this

nature are found in countries lacking good vital registration systems. They also reflect sources of error (measurement, faulty sample frames, and non-response), which are beyond the scope of assessing sample uncertainty and stochastic uncertainty. Given this, we believe that our method offers a low-cost and efficient means of assessing the stochastic uncertainty in sample based IMR estimates, thereby serving as a validity check when these estimates are used to make policy decisions. Swanson and Tayman [22, p. 304] suggest these characteristics are important components in deciding what methods to use in developing estimates. Supporting this concluding statement is the evidence from the validity test. For Ghana, our results suggest that the sample based estimates are reliable for all 13 sample regions. This includes the regions that have relative large standard errors (over 8.0) in terms of the sample sizes, Ashanti, Upper East, and Volta.

While the beta-binomial model has been used in medical research [28–30], consumer studies [31], bioinformatics [32] and public health research [33,34], it has not found much traction in economic, demographic, and sociological approaches to the study of population. This is surprising on two counts: (1) the components of demographic change, births, deaths, and migration, can all be constructed as rates that are inherently binomial variables; and (2) the method is simple to use, explain, and understand.³ This paper illustrates one such use with a sub-set of the mortality component, the infant mortality rate. Although we focus on an African application, the method can be applied to many other situations where small, inherently binomial, numbers are present and affected by stochastic uncertainty, whether they are from a scientific sample, as is the case in this paper, or a “complete count” as found in a reliable vital registration system [17].

References

- [1] Kitagawa E., and Hauser P. (1973). *Differential Mortality in the United States: A Study in Socio-economic Epidemiology*. Cambridge: Harvard University Press.
- [2] Link W., and Hahn D. (1996). Empirical bayes estimation of proportions with application to cow-bird parasitism rates. *Ecology* 77(8): 2528-2537.

³Although Green and Armstrong [38] discuss simple vs. complex methods in terms of forecasting, their discussion applies here in that the beta-binomial approach falls into the simple methodological category rather than the complex category. Adapting their discussion to methods in general, the work of Green and Armstrong [38] suggests that while there is no evidence that shows complexity improves accuracy, complexity remains popular among: (1) researchers, because they are rewarded for publishing in highly ranked journals, which favor complexity; (2) methodologists, because complex methods can be used to provide information that support decision makers’ plans; and (3) clients, who may be reassured by incomprehensibility. We believe that the argument by Green and Armstrong [38] can be applied to Bayesian methods, which represents the “complex” alternative to the “simple” Beta-binomial approach. We prefer the Beta-binomial approach, however, not only because of the argument presented by Green and Armstrong, but also because the application of a Bayesian approach can be difficult, effortful, opaque and even counter-intuitive [39].

- [3] Stockwell E., Bedard M., Swanson D.A., and Wicks J. (1987). Public policy and the socioeconomic mortality differential in infancy. *Population Research and Policy Review* 6(Fall): 105-121.
- [4] Stockwell E., Goza F., and Balisteri K. (2005). Infant mortality and socioeconomic status: new bottle, same old wine. *Population Research and Policy Review* 24: 387-39.
- [5] Hummer R. (2005). Income, race, and infant mortality: comment on Stockwell et al. *Population Research and Policy Review* 24: 405-409.
- [6] World Health Organization (2018). Global Health Observatory (GHO) Data. Available online at http://www.who.int/gho/child_health/mortality/neonatal_infant_text/en/ Accessed February 5, 2018.
- [7] World Health Organization (2018). Global Health Observatory (GHO) Data. Available online at <http://apps.who.int/gho/data/node.country.country-SLE> Accessed February 5, 2018.
- [8] UNICEF (2015). Newborn Mortality. Available online at https://www.unicef.org/ghana/REALLY_SIMPLE_STATS_-_Issue_1.pdf (Accessed 26 March 2018).
- [9] Ghana Statistical Service (GSS), Ghana Health Service (GHS), and ICF International (2015). *Ghana Demographic and Health Survey 2014*. Rockville, Maryland, USA: GSS, GHS, and ICF International.
- [10] Foote K., and Hill A. (1993). Trends in childhood mortality. In: Foote K., Hill K., and Martin L. (Eds). *Demographic change in Sub-Saharan Africa*. National Academy of Sciences. Washington, DC. National Academies Press. pp. 153-217.
- [11] Ghana Statistical Service (2018). Surveys. At <http://www.statsghana.gov.gh/surveys.html>.
- [12] Ghana Statistical Service (GSS), Ghana Health Service (GHS), and ICF International (2009). *Ghana Demographic and Health Survey 2008*. Rockville, Maryland, USA: GSS, GHS, and ICF International.
- [13] Ghana Statistical Service (GSS), Noguchi Memorial Institute for Medical Research (NMIMR), and ORC Macro (2004). *Ghana Demographic and Health Survey 2003*. Calverton, Maryland, USA: GSS, NMIMR, and ORC Macro.
- [14] Ghana Statistical Service (GSS), Institute for Resource Development/Macro Systems (IRC) (1989). *Ghana Demographic and Health Survey 1988*. Accra: Ghana and Columbia, Md: GSS and IRC. Md.: Institute for Resource Development/Macro Systems, Inc.
- [15] World Bank (2018). Infant Mortality Rates by Country and Year, 1960 to 2016, file = API_SP.DYN.IMRT.IN_DS2_en_excel_v2.xls, available online at <https://data.worldbank.org/indicator/SP.DYN.IMRT.IN>, where the excel file can be downloaded by going to the "download" tab and clicking on EXCEL.
- [16] Link B., and Phelan J. (1995). Social conditions as fundamental causes of disease. *Journal of Health and Social Behavior* (extra issue): 80-98.
- [17] Swanson D.A. (forthcoming). Estimating the underlying infant mortality rates for small populations, even those reporting zero infant deaths: A case study of 66 local health areas in British Columbia. *Canadian Studies in Population*.
- [18] Doob J. (1953). *Stochastic Processes*. New York. John Wiley & Sons.
- [19] Ramsey M., and Ellison S. (eds.) (2007). *Eurachem/EUROLAB/CITAC/Nordtest/AMC Guide: Measurement uncertainty arising from sampling: a guide to methods and approaches*. Torino, Italy. Eurachem Secretariat.
- [20] Karlin S., and Taylor S. (2001). *A First Course in Stochastic Processes. 2nd edition*. San Diego: Academic Press.
- [21] Graham C., and Talay D. (2013). *Stochastic Simulation and Monte Carlo Methods*. New York: Springer.
- [22] Swanson D.A., and Tayman J. (2012). *Subnational Population Estimates*. Dordrecht, The Netherlands: Springer.
- [23] Brass W., Coale A., Demeny P., Heisel D., Lorimer F., Romaniuk A., and Van de Walle E. (1968). *The Demography of Tropical Africa*. Princeton: Princeton University Press.
- [24] United Nations. (1967). *Manual IV, methods of estimating basic demographic measures from incomplete data*. New York, NY: United Nations.

- [25] Baker J., Alcantara A., and Ruan X.M. (2011). A stochastic version of the brass PF ratio adjustment of age-specific fertility schedules. *PLoS One*. 6(8): e23222.
- [26] Moultrie T., Dorrington R., Hill A., Hill K., Timæus I., and Zaba B. (2013). *Tools for Demographic Estimation*. Paris: International Union for the Scientific Study of Population.
- [27] Siegel J., and Swanson D. (2004). *The methods and materials of demography, 2nd Edition*. Los Angeles: Academic/Elsevier Press.
- [28] Kim J., and Lee J. (2013). Simultaneous confidence intervals for a success probability and intraclass correlation, with an application to screening mammography. *Biometrical Journal*. 55(6): 944-954.
- [29] Arostegui I., Nuñez-Antón V., and Quintana J. (2007). Analysis of the Short Form-36 (SF-36): the beta binomial distribution approach. *Statistics in Medicine*. 26: 1318-1342.
- [30] Young-Xu Y., and Chan K. (2008). Pooling overdispersed binomial data to estimate event rate BMC Medical Research Methodology 8:58.
- [31] Chatfield C., and Goodhardt G. (1970). The beta-binomial model for consumer purchasing behavior. *Applied Statistics*. 19: 240-250.
- [32] Pham T., Piersma S., Warmoes M., and Jimenez C. (2010). On the beta-binomial model for analysis of spectral count data in label-free tandem mass spectrometry-based proteomics. *Bioinformatics*. 2(3): 363-369.
- [33] Alanko T., and Lemmens P. (1996). Response effects in consumption surveys: an application of the Beta-binomial model to self-reported drinking frequencies. *Journal of Official Statistics*. 12(3): 253-273.
- [34] Gakidou E., and King G. (2002). Measuring total health inequality: adding individual variation to group-level differences. *International Journal of Equity in Health* 1:3 doi: 10.1186/1475-9276-1-3.
- [35] Murray C. (1996). Rethinking DALYs. In: Murray C., and Lopez A. (eds.). *The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020*. Cambridge, MA: Harvard School of Public Health. pp. 1-98.
- [36] Reidpath D., and Allotey P. (2003). Infant mortality rate as an indicator of population health. *Journal of Epidemiology and Community Health*. 57: 344-346.
- [37] Guillot M., Lim S., Torgasheva L., and Denisenko M. (2013). Infant mortality in Kyrgyzstan before and after the break-up of the Soviet Union. *Population Studies*. 67(3): 335-352.
- [38] Green K., and Armstrong J.S. (2015). Simple versus complex forecasting: the evidence. *Journal of Business Research*. 68: 1678-1685.
- [39] Goodwin P. (2015). When simple alternatives to bayes formula work well: reducing the cognitive load when updating probability forecasts. *Journal of Business Research*. 68: 1686-1691.