Research Article

**Child mortality levels and trends:
A new compositional approach**

Fatine Ezbakhe

Agustí Pérez-Foguet

Contents

Child mortality levels and trends: A new compositional approach

Fatine Ezbakhe[1,2]

Agustí Pérez-Foguet[1]

Abstract

BACKGROUND
Trend analysis of child mortality is vital to evaluate countries' progress towards achieving the Sustainable Development Goal on health (SDG 3). However, strictly speaking, child mortality data are probabilities, and thus subject to non-negativity and constant-sum constraints.

OBJECTIVE
Our objective is to assess the application of compositional data analysis for estimating levels and trends in child mortality.

METHODS
We compare two data transformations: logit, which is widely used in child mortality estimation, and isometric log-ratio (ILR), which is specifically designed for compositional data. We use publicly available household survey data on neonatal (NMR) and under-five (U5MR) mortality ratios in sub-Saharan Africa.

RESULTS
Although both data transformations yield similar estimates, only the ILR transformation is consistent with the compositional properties of child mortality data. However, the ILR suffers from one key drawback: it requires complete data series, with pairs of observations for both NMR and U5MR. As a result, ILR entails excluding a large amount of available data from the regression analysis.

CONCLUSIONS
Complete data is needed to be able to undertake a compositional trend analysis of child mortality. This gap in data can be closed by employing imputation strategies that replace missing values in the existing datasets, and by developing new methods for the indirect estimation of NMR from summary birth history data, as it is currently done for U5MR.

---

[1] Engineering Sciences and Global Development (EScGD), Department of Civil and Environmental Engineering (DECA), Barcelona School of Civil Engineering, Universitat Politècnica de Catalunya, Barcelona, Spain.
[2] Corresponding author: fatine.ezbakhe@unige.ch.

CONTRIBUTION
This paper extends the literature on child mortality estimation by examining the application of compositional data analysis to this field. It constitutes a first step towards building a Bayesian compositional regression approach for child mortality estimation.

1. Introduction

The ongoing decline in child mortality is considered one of the most critical successes in public and population health of the past three decades. The deaths of children under 5 years old have fallen from 12.5 million per year in 1990 to 5.3 million per year in 2018, even as the world's population under age 5 grew by nearly 32.7 million (UNICEF 2019; UNESA 2019). Notwithstanding this progress, there is still a heavy burden of child deaths due to preventable or treatable causes such as pneumonia, malaria, and diarrhoea. Such a burden has both social and economic consequences: in the WHO African region alone,[3] the cost of child mortality amounted to 150.3 billion US dollars in 2013 (i.e., approximately 6% of the combined GDP in the region) (Kirigia et al. 2015). In recognition of the crucial need to further combat child mortality, the third Sustainable Development Goal (SDG 3) of the 2030 Agenda explicitly calls for countries to "ensure healthy lives and promote wellbeing for all at all ages" (UNGA 2015). In particular, Target 3.2 specifies the end of preventable deaths of newborns and children under 5 by lowering the neonatal and under-5 mortality rates to at least 12 and 25 deaths per 1,000 live births, respectively, by 2030.

Achieving this ambitious child survival target goes beyond ensuring universal access to adequate, good-quality, and affordable healthcare for women and children. It also requires understanding the levels and trends in child mortality in order to evaluate countries' performances and identify effective policies (UNICEF 2019). That is why measuring and monitoring child mortality is a global priority. However, tracking progress towards reducing child mortality can be challenging, particularly in developing countries with dysfunctional vital registration systems. According to Mahapatra et al. (2007), vital statistics are unavailable or of poor quality in 111 countries, mainly in sub-Saharan Africa, South-East Asia, and the Western Pacific, which represent 72% of the world's population. This lack of reliable data inevitably takes a toll on the effectiveness of public health policymaking.

To overcome the absence of reliable vital registration data in many countries, the United Nations Inter-agency Group for Child Mortality Estimation (UN IGME) produces

---

[3] The World Health Organization (WHO) divides the world into six WHO regions for reporting and analysis. The WHO African region comprises 47 countries, listed at https://www.who.int/about/regions/afro/en/.

and publishes estimates of child and young adolescent mortality rates every year (UN IGME 2018). The UN IGME provides child mortality estimates for three different age intervals: (1) neonatal mortality rate (NMR), i.e., the number of deaths within the first 28 days of life per 1,000 live births; (2) the infant mortality rate (IMR), i.e., the number of deaths among children under the age of 1 year per 1,000 live births; and (3) under-5 mortality rate (U5MR), i.e., the number of deaths of children up to the age of 5 per 1,000 live births. With this input data, the UN IGME generates child mortality estimates for years of interest using a Bayesian B-splines Bias-adjusted (B3) regression model (Alkema and New 2014; Alkema et al. 2014a; Alexander and Alkema 2018). Several alternative curve-fitting models have been developed both at national (Rajaratnam et al. 2010; Hill et al. 2012; Wang et al. 2014; Alkema et al. 2014b) and subnational levels (Dwyer-Lindgren et al. 2014; Mercer et al. 2015; Pezzulo et al. 2016; Golding et al. 2017; Chao et al. 2018; Wakefield et al. 2018; Li et al. 2019). The UN IGME's B3 model also adjusts the errors and biases in the data.

However, besides accounting for the inherent uncertainty in child mortality data, the analysis should also consider its compositional nature. Strictly speaking, child mortality indicators are not rates but probabilities calculated according to the conventional life-table approach (Rutstein 1984) and are thus naturally constrained. Indeed, the sum of probabilities of dying in the neonatal (0–28 days), post-neonatal (29–364 days), and childhood (1–4 years) age intervals and the probability of surviving beyond 5 years must equal 1. This constant sum constraint makes it impossible to follow the usual Euclidean geometry, since data belongs to a subspace of the Euclidean space, known as the Aitchison simplex, with its geometrical structure and operations (Aitchison 1982; Pawlowsky-Glahn, Egozcue, and Tolosana-Delgado 2015). Therefore, child mortality rates must not be analyzed separately, as this may lead to spurious correlations and, consequently, to wrong interpretations. In the B3 model, the compositional nature of data is accounted for to some extent by considering logarithms and ratios. For instance, the UN IGME models the U5MR in the log scale (i.e., $\log(\text{U5MR})$). For the IMR, it considers the log-odds transformation of the ratio r between the IMR and the median B3 model estimates of U5MR (i.e., $\log(r/(1-r))$). For the NMR, it considers the ratio between NMR and the difference between U5MR and NMR (i.e., $\text{NMR}/(\text{U5MR}-\text{NMR})$). However, as Aitchison (1999) emphasized, to be fully in line with the principles of compositional data analysis, log-ratio transformations between the compositional parts are needed.

The application of compositional analysis to mortality data is not new. Oeppen (2008) explores the use of centered log-ratio transformation for forecasting mortality by cause of death. Similarly, Salomon and Murray (2001) develop a compositional model based on additive log-ratios to predict cause-of-death patterns by age and sex, and Bergeron-Boucher et al. (2017, 2018) apply centered log-ratios to forecast the

distribution of deaths of subpopulations coherently. Other researchers focus on using isometric log-ratios to model trends in other SDG-related indicators, such as safe water and sanitation (Pérez-Foguet, Giné-Garriga, and Ortego 2017; Ezbakhe and Pérez-Foguet 2019), clean energy (Marcillo-Delgado, Ortego, and Pérez-Foguet 2019), and health-related outcomes (Carson et al. 2016; Fairclough et al. 2017). However, the modelling of compositional trends in child mortality remains an unexplored area.

In that context, this paper aims to assess the application of compositional data analysis for estimating levels and trends in child mortality. Specifically, our intention is twofold:

- From a theoretical perspective, we investigate the need to consider the compositional properties of child mortality data – in particular their unit-sum constraint – by comparing two data transformations based on logarithms of ratios: (1) the log of the odds (LOGIT), and (2) the isometric log-ratio (ILR).
- From a practical viewpoint, we examine the suitability of a compositional approach to child mortality modelling by comparing the ILR-based estimates with those provided by the UN IGME. Although these models use two significantly different regression methods – Generalized Additive Model (GAM) and Bayesian penalized B-splines, respectively – we complete this comparison as an initial step towards building a Bayesian compositional regression approach to child mortality estimation.

To this end, we use all publicly available household survey data on the child mortality indicators used in SDG 3 monitoring – neonatal (NMR) and under-5 (U5MR) mortality rates – in sub-Saharan Africa. We select this region for two reasons. First, it accounted for nearly 52.5% of global under-5 deaths in 2018 (UNICEF 2019). Second, it fully overlaps with the African countries included in UNICEF's EQUItable Impact Sensitive Tool.[4]

The remainder of the paper is structured as follows. Section 2 provides the background to the use of household survey data in child mortality monitoring (2.1) and the general principles behind compositional data analysis (2.2). Section 3 presents an overview of the method (3.1) and data (3.2) used in our analyses. Section 4 presents and discusses the results from applying the two data transformations – LOGIT and ILR – to child mortality (4.1), by comparing our ILR-based estimates and those provided by the UN IGME (4.2), and by analyzing the geographical distribution of child mortality in sub-

---

[4] The EQUitable Impact Sensitive Tool (EQUIST) is a medium-term analysis and strategic planning tool that aims to help decision-makers maximize the effect of public policies and improve health and nutrition for children from low-income countries (UNICEF 2017).

Saharan Africa (4.3). Section 5 concludes the paper by summarizing the primary outcomes of our analyses and indicating directions for further research.

2. Background

In this section, we first provide a background on the monitoring of child mortality, in particular the use of household surveys (2.1). We then explain the statistical methodology for compositional data analysis (2.2).

2.1 Child mortality monitoring

The responsibility for monitoring and assessing child mortality at the global, regional, and country level lies with the United Nations Children's Fund (UNICEF). Together with other members[5] of the UN Inter-agency Group for Child Mortality Estimation (IGME), UNICEF estimates child mortality every year to monitor progress and shares it in their public database (http://www.childmortality.org). The UN IGME first reviews and compiles all available nationally representative data relevant to the estimation of child mortality, including data from civil registration systems, population censuses, and household surveys, and assesses their quality to exclude those with substantial errors.

The most reliable data sources for child mortality estimates are civil registration and vital statistic (CRVS) systems, in which all births and deaths are routinely registered and certified. Unfortunately, most developing countries lack comprehensive CRVS systems. In the absence of continuous recording systems, measures of child mortality are derived from alternative data sources, most notably periodic, nationally representative household surveys (Hill et al. 2015), using both direct and indirect methods. Direct estimation approaches collect child mortality from the full birth histories (FBHs) of women of reproductive age (i.e., 15 to 49 years old). In an FBH, women report the date of birth, sex, survival status, age (if alive), and age at death (if dead) for each of their births. Probabilities of dying in childhood are then computed based on synthetic cohort life tables (Rutstein and Rojas 2006). However, this approach is time-consuming and expensive due to the extensive questionnaires and training of interviewers. Indirect estimation methods, on the other hand, use summary birth histories (SBHs), whereby women only report the total number of children they have given birth to and the number who have died – or equivalently the number still alive – at the time of the survey. Instead

---

[5] The other members of the UN IGME are the World Health Organization (WHO), the World Bank, and the United Nations Population Division of the Department of Economic and Social Affairs (UNESA).

of a full distribution of births and deaths over time as in FBHs, SBHs only provide the proportions of children dead at the time of the survey. In SBHs, probabilities of dying in childhood are derived from modelling the relationship between the proportions of children dead and the age of the women (Brass 1971; Zlotnik and Hill 1981).

Most countries turn to household surveys to collect data on child mortality. As seen in Table 1, household surveys account for 91.1% and 79.4% of the total number of data series compiled for NMR and U5MR, respectively, which represent 40.2% and 55.7% of all mortality observations in the database. Amongst the most common household surveys are Demographic and Health Surveys (DHS), providing 59.8% and 52.4% of household data in NMR and U5MR, respectively.

Table 1: **Data availability for neonatal (NMR) and under-5 (U5MR) mortality rates from the UN IGME public database (extracted on 28 May 2019)**

Indicator	Type	Source	Number of countries	Number of series	Number of observations	% excluded by UN IGME	% with unreported standard errors
NMR	Vital registration	VR	115	20	3,320	29.2	-
		SVR	4	4	91	1.1	-
	Censuses	CEN	0	0	0	0	-
	Household surveys	MICS	37	22	222	27.0	16.7
		Other MICS	0	0	0	0.0	0.0
		DHS	90	73	1,372	8.1	4.6
		Other DHS	72	49	362	14.1	21.5
		LSMS	0	0	0	0	0
		Other surveys	59	103	337	15.4	47.2
U5MR	Vital registration	VR	136	78	5,619	42.5	-
		SVR	4	7	140	21.4	-
	Censuses	CEN	137	128	2,332	48.1	-
	Household surveys	MICS	79	58	1,008	44.5	21.4
		Other MICS	4	3	24	16.7	100.0
		DHS	90	176	5,333	40.1	22.5
		Other DHS	96	123	1,465	39.9	37.8
		LSMS	1	1	1	100.0	100.0
		Other surveys	130	462	2,349	41.6	88.7

Notes: The acronyms are as follows: VR (Vital Registration), SVR (Sample Vital Registration), CEN (Census), MICS (Multiple Indicator Cluster Survey), DHS (Demographic and Health Survey) and LSMS (Living Standard Measurement Survey). 'Other MICS' category includes National MICS, and 'Other DHS' includes Interim DHS, Special DHS, National DHS, World Fertility Survey, Malaria Indicator Survey, and AIDS Indicator Survey.

We highlight two other relevant features from the child mortality database. First, the UN IGME excludes a significant number of data points from their estimation process (an average of 10.5% and 43.9% for NMR and U5MR, respectively), because of their substantial degree of non-sampling errors or omissions. Second, information on sampling

errors in NMR observations is missing in more than 15% of the household surveys, and nearly 62% in U5MR observations.

Non-sampling errors may arise due to many different factors, including non-coverage, non-response, questionnaires of inadequate quality, or poor survey implementation and data collection and processing (Lesser and Kalsbeek 1999). Specifically, errors in data collection can be due to the underreporting of deceased children (especially of neonatal deaths) or misreporting of ages at death (in particular age heaping around age 1) (Guillot et al. 2012).

While non-sampling errors can be minimized in many ways (e.g., proper design of survey questionnaire and data collection), sampling errors will always exist, as the sample size is always smaller than the size of the population. Such sampling errors can be quite significant in child mortality estimates. A review by Korenromp et al. (2004) of sampling errors from Demographic and Health Surveys (DHS) in various sub-Saharan African countries reveals median relative errors of 5.6% and 4.4% for IMR and U5MR, respectively. This considerable amount of uncertainty is mainly because "most household surveys are not designed to produce highly accurate estimates of child mortality, but rather aim for high accuracy of a number of other indicators" (UNESA 2011). For instance, in the Multiple Indicator Cluster Surveys (MICS), child mortality rates are not selected as key indicators on which to base the calculation of the sample size. This is because the sizes that would be required to measure child mortality indicators with the same precision as recommended for other indicators would be too large and impractical. Other indicators, such as school attendance or immunization coverage, are used instead (UNICEF 2006).

That is why, in addition to further minimizing these errors, uncertainty assessment of the estimates – in the form of uncertainty intervals, for example – is indispensable for an evidence-based analysis of child mortality levels and trends. Failure to conduct and report such uncertainty intervals may lead to misinterpretation of rates and trends, and ultimately undermine effective policymaking for child mortality reduction.

2.2 Compositional data analysis

Compositional data are non-negative multivariate data that are some part of a whole. They are usually recorded in closed form, summing to a constant (e.g., percentages summing to 100% or proportions summing to 1). Such data have particular and essential properties that arise from the fact that they represent parts of some whole (Pawlowsky-Glahn and Egozcue 2006): they are a vector of strictly positive real numbers with a constant sum constraint:

$$\boldsymbol{x} = (x_1, x_2, \ldots, x_D); \; x_i > 0 \qquad i = 1, 2, \ldots, D; \; \sum_{i=1}^{D} x_i = \kappa \qquad (1)$$

The elements of a composition, x_i, are called components or parts, and the only relevant information is contained in the ratios between them. This feature conditions the relationships between the components. For instance, if one component is decreasing over time, at least one other component has to increase to preserve the constant sum. As a result, compositional data are enclosed in a subspace where they can only vary between 0 and the radix value (κ). This subspace, known as the simplex, does not follow the rules of Euclidean geometry, which makes the use of standard statistical techniques inappropriate for the analysis of compositional data (Aitchison 1999). Applying standard statistical approaches to 'raw' compositional data might lead to spurious correlations and inferences.

Thus, because of this particular geometry, working in the simplex can be counterintuitive. As an alternative, compositional data may be transformed into the real space where classic statistics can be applied (Pawlowsky-Glahn, Egozcue, and Tolosana-Delgado 2015). These transformations are based on log ratios between components and lead to 'open' data, called coordinates, which can take any real value between $-\infty$ and ∞. Several log-transformations have been proposed, including the additive log-ratio (ALR), the centred log-ratio (CLR), and the isometric log-ratio (ILR) (Aitchison 1982; Egozcue et al. 2003).

In this paper we use the ILR transformation, which represents the composition given a particular orthonormal basis in the simplex:

$$\boldsymbol{z} = \mathrm{ilr}(\boldsymbol{x}) = \log(\boldsymbol{x}) \cdot \boldsymbol{V} \qquad (2)$$

where $\boldsymbol{x}$ is the vector with the D parts of the composition, $\boldsymbol{V}$ is a $D \cdot (D-1)$ matrix representing an orthonormal basis in the simplex, and $\boldsymbol{z}$ is the resulting vector with $D-1$ coordinates of the composition in that basis $\boldsymbol{V}$.

There are several ways to define orthonormal bases in the simplex, one of which consists of a sequential binary partition (SBP) of the composition (Pawlowsky-Glahn, Egozcue, and Tolosana-Delgado 2015). An SBP represents a hierarchy of the parts of a composition and contains successive splits of the parts into two non-overlapping groups, coded by the signs + and –, respectively. The coordinates of the composition in the orthonormal basis $\boldsymbol{V}$ can be obtained from the SBP as:

$$z_i = \sqrt{\frac{r_i s_i}{r_i + s_i}} \log \left(\frac{(\Pi_+ x_{ij})^{1/r_i}}{(\Pi_- x_{ik})^{1/s_i}} \right); \qquad i = 1, 2, \ldots, D-1 \qquad (3)$$

where z_i is the ith coordinate (or balance) of the composition, x_{ij} and x_{ik} are the components coded as + and − in the ith partition, and r_i and s_i are the number of parts with positive and negative signs in that partition, respectively.

Once the data are transformed into ILR balances, standard statistical approaches can be applied. Finally, regression points can be back-transformed to the original space using the inverse ILR:

$$\boldsymbol{x} = \mathrm{ilr}^{-1}(\boldsymbol{z}) = \mathcal{C}[\exp(\boldsymbol{V} \cdot \boldsymbol{z})] \tag{4}$$

where $\boldsymbol{z}$ contains the ILR coordinates of $\boldsymbol{x}$ with respect to the basis $\boldsymbol{V}$, and $\mathcal{C}$ is the closure operator:

$$\mathcal{C}[\boldsymbol{x}] = \left(\frac{x_1}{\sum_{i=1}^{D} x_i}, \frac{x_2}{\sum_{i=1}^{D} x_i}, \dots, \frac{x_D}{\sum_{i=1}^{D} x_i} \right); \quad i = 1, 2, \dots, D. \tag{5}$$

3. Methods overview

In this section, we first describe our method for the compositionally based estimation of child mortality (3.1). We then present the data used to estimate child mortality in sub-Saharan countries (3.2).

3.1 Proposed approach

We estimate the NMR and U5MR for each country during 1990–2018 – or earlier if data is available – using a Generalized Additive Model (GAM). We choose the GAM for three main reasons. First, as a generalized linear model, the GAM uses a link function to establish a relationship between the mean of the response variable and a smoothed function of the set of explanatory variables. This limits the error in the prediction of the response variable from various probability distributions. In our model, we apply thin-plate regression splines with four degrees of freedom for the smoothing functions (Wood 2003). Second, the GAM allows for a non-parametric regression, where there is no underlying assumption of linearity between the response variable and the covariates. Thus, the GAM can deal with non-linear and non-monotonic relationships between the response and the explanatory variables. Third, the GAM has been widely used to explore demographic and health data, in particular child mortality data (Ayele, Zewotir, and Mwambi 2015; Dlamini, Melesse, and Mwambi 2017; Burstein et al. 2018).

We consider child mortality data as 3-part compositions, $\boldsymbol{x} = (x_1, x_2, x_3)$, where x_1 is the probability of dying in the neonatal age interval (0 to 28 days), x_2 the probability of dying in the post-neonatal and childhood age intervals (29 days to 4 years), and x_3 the probability of surviving beyond 5 years:

$$x_1 = \text{NMR}/1000; \, x_2 = (\text{U5MR} - \text{NMR})/1000; \, x_3 = 1 - \text{U5MR}/1000. \qquad (6)$$

We perform two data transformations:
- Log of the odds (LOGIT) transformation, where we fit the model to the logarithm of the odds, as in Equation 7.

$$z_i = \log\left(\frac{x_i}{1-x_i}\right), \, i = 1,2,3 \qquad (7)$$

- Isometric log-ratio (ILR) transformation, in which we fit the model to the $D - 1$ balances, as in Equation 3.

For ILR, we create an SBP[6] that mimics the $\text{NMR} \cdot (\text{U5MR} - \text{NMR})$ ratio modelled by the UN IGME:

Balance (z_i)	x_1	x_2	x_3	r	s
1	+1	+1	−1	2	1
2	+1	−1	0	1	1

With this established SBP, the orthonormal basis is:

$$\boldsymbol{V} = \begin{bmatrix} \sqrt{1/6} & \sqrt{1/2} \\ \sqrt{1/6} & -\sqrt{1/2} \\ -\sqrt{2/3} & 0 \end{bmatrix}$$

The ILR coordinates can be computed, following Equation 3, as:

$$z_1 = \sqrt{\frac{2}{3}}\log\left(\frac{\sqrt{x_1 x_2}}{x_3}\right) = \sqrt{\frac{2}{3}}\log\left(\frac{\sqrt{\text{NMR} \cdot (\text{U5MR} - \text{NMR})}}{1000 - \text{U5MR}}\right) \qquad (8)$$

$$z_2 = \frac{1}{\sqrt{2}}\log\left(\frac{x_1}{x_2}\right) = \frac{1}{\sqrt{2}}\log\left(\frac{\text{NMR}}{\text{U5MR} - \text{NMR}}\right) \qquad (9)$$

---

[6] Although the actual basis used to compute the ILR coordinates is irrelevant (i.e., the back-transformed results are the same whichever basis), it is helpful to choose a basis that allows the interpretation of results on the level of coordinates.

The regression estimates are back-transformed to the original space using the inverse LOGIT function (i.e., $e^{z_i}/(1 + e^{z_i})$) and the inverse ILR shown in Equation 4. The NMR and U5MR estimates are finally derived from the values of x_1 and x_2 (as in Equation 6).

Similar to previous studies (Bermejo et al. 2015; Minnery et al. 2015; Hodge et al. 2014), we construct confidence intervals of mortality rates via simulation techniques. This involves generating 1,000 simulations of the survival probability for each age group – neonatal and under-5 – assuming a Binomial distribution, $B(p, n)$. [7] The survival probability is computed as $1 - \text{MR}$, and the sample size n is derived from the standard errors of rates (i.e., $se = \sqrt{\text{MR}(1 - \text{MR})/n}$), where MR is the mortality rate (i.e., NMR/1000 and U5MR/1000, respectively). Finally, the 90% confidence intervals are obtained from the 5th and 95th percentiles of the simulations.

The R code of our approach is publicly available and downloadable (Ezbakhe and Pérez-Foguet 2019), and uses the packages *Compositions* (van den Boogaart, Tolosana-Delgado, and Bren 2018) and *Gam* (Hastie 2018).

3.2 Child mortality data

In our analysis, we only consider data from censuses and household survey series, and those series deemed of good quality by the UN IGME. For data with unreported sampling standard errors we impute an error of 2.5% for census observations and 10% for household surveys, as done by Alkema and New (2014). On the other hand, since we consider data as 3-part compositions, only year series with observations for both rates, NMR and U5RM, are included in the analysis, which significantly impacts the amount of data incorporated into our modelling procedure. There is notably less data available for NMR than U5MR (i.e., 247 year series with 2,293 observations for U5MR versus 823 series with 10,180 for NMR).

The number of compositional data points considered for each of the 48 countries of sub-Saharan Africa are shown in Table 2. On average, only 18% of the observations include both NMR and U5MR. A clear example of this is Senegal: from the 44 and 73 time series for NMR and U5MR, respectively, only 34 include information for both rates (resulting in 41 observations instead of the 186 available for U5MR). Furthermore, there are 9 countries with less than 4 observations for both NMR and U5MR – Central African Republic, Comoros, Djibouti, Equatorial Guinea, Gabon, Gambia, Seychelles, Sierra Leone, and South Sudan – that we exclude from the analysis because of their lack of sufficient data.

---

[7] We use a Binomial distribution to model child mortality because it is the preferred distribution for dealing with counts (in this case, the number of child deaths).

Table 2: **Data availability for neonatal (NMR) and under-5 (U5MR) mortality rates from household surveys and censuses in sub-Saharan African countries (from the UN IGME public database, extracted on 28 May 2019)**

Country		NMR		U5MR		BOTH	
		Years	Points	Years	Points	Years	Points
AGO	Angola	5	5	23	24	5	5
BEN	Benin	15	25	65	105	14	21
BWA	Botswana	9	9	28	28	4	4
BFA	Burkina Faso	17	23	67	110	15	19
BDI	Burundi	15	15	49	50	5	5
CPV	Cape Verde	8	8	16	16	5	5
CMR	Cameroon	25	30	68	92	19	20
CAF	Central African Republic	5	5	32	32	3	3
TCD	Chad	12	15	49	57	8	8
COM	Comoros	5	5	10	10	0	0
COG	Congo	13	13	28	28	6	6
CIV	Ivory Coast	25	25	68	83	14	14
COD	Democratic Republic of Congo	10	10	28	28	10	10
DJI	Djibouti	6	6	16	16	1	1
GNQ	Equatorial Guinea	3	3	10	10	3	3
ERI	Eritrea	13	13	30	34	4	4
ETH	Ethiopia	12	20	47	77	6	10
GAB	Gabon	6	6	8	8	2	2
GMB	Gambia	0	0	27	28	0	0
GHA	Ghana	26	43	88	115	11	12
GIN	Guinea	21	21	58	78	10	10
GNB	Guinea-Bissau	10	10	18	18	10	10
KEN	Kenya	19	35	87	127	13	21
LSO	Lesotho	12	20	55	58	5	5
LBR	Liberia	15	20	57	63	4	4
MDG	Madagascar	12	20	49	68	7	10
MWI	Malawi	28	35	82	164	24	25
MLI	Mali	16	20	41	94	16	20
MRT	Mauritania	17	20	52	71	12	12
MOZ	Mozambique	11	15	44	76	10	14
NAM	Namibia	20	20	26	29	5	5
NER	Niger	23	25	45	88	20	20
NGA	Nigeria	16	25	45	75	7	15
RWA	Rwanda	21	29	60	108	17	23
STP	Sao Tome and Principe	10	10	24	24	10	10
SEN	Senegal	44	60	73	186	34	41
SYC	Seychelles	0	0	10	10	0	0
SLE	Sierra Leone	5	5	68	69	3	3
SOM	Somalia	5	5	10	10	5	5

Table 2: **(Continued)**

Country		NMR		U5MR		BOTH	
		Years	Points	Years	Points	Years	Points
ZAF	South Africa	10	10	24	24	7	7
SSD	South Sudan	1	1	13	13	1	1
SDN	Sudan	17	17	88	93	6	6
SWZ	Swaziland	15	15	34	36	8	8
TGO	Togo	10	15	58	70	5	7
UGA	Uganda	23	35	68	126	13	19

Note: "Years" = the number of year series; "Points" = the number of data observations.

4. Results and discussion

4.1 Comparative analysis 1: LOGIT vs. ILR data transformations

Figures 1 and 2 show the comparison of the child mortality estimates obtained with the two data transformations, LOGIT and ILR, for Malawi and Mauritania.

The LOGIT data transformation provides three univariate coordinates, each representing separately the log-odds of the mortalities and survival probabilities. In the case of Malawi, the log-odds for mortality under 1 month old (z_1) and mortality from ages 1 to 5 (z_2) display decreasing values (between 1970 and 2018, z_1 and z_2 declined by 1.3 and 1.9 points, respectively); whereas log-odds for survival beyond age 5 (z_3) increase by 1.9 points. These figures indicate both a reduction in child mortality and improved life expectancy over time. Mauritania presents a different trend in the LOGIT coordinates: z_1 first increases between 1968 and 1975 (by 0.4 points) and then decreases until 2018 (by 0.5 points) and z_2 starts with a decline of 0.7 points from 1968 to 1989 and increases 0.8 points afterwards, while z_3 shows an acceleration of 0.4 points between 1968 and 1992 and stabilizes until 2018. These smaller changes in all z_1, z_2, and z_3 translate into weaker progress in child mortality reduction.

Figure 1: **Mortality estimates for Malawi, with (a) logit and (b) isometric log-ratio transformations**
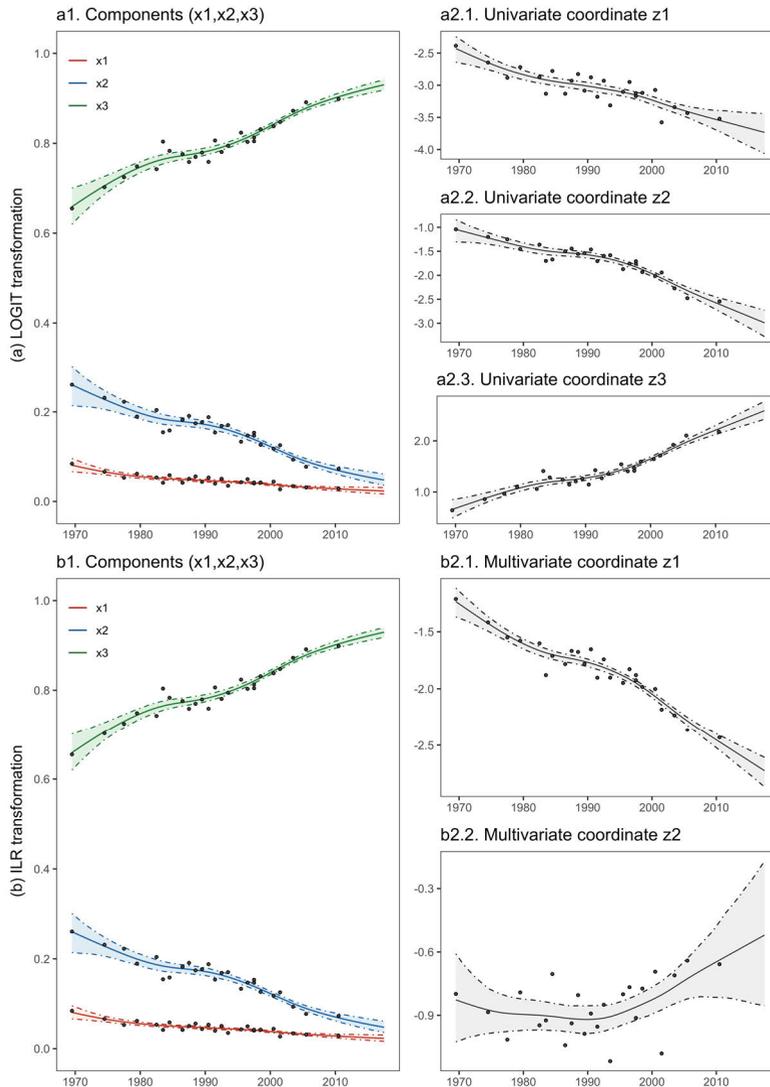
Notes: The regression results in the coordinates are In black: the three univariate coordinates for LOGIT (in a2.1, a2.2 and a2.3) and the two multivariate coordinates for ILR (in b2.1 and b2.2). In colour (a1 and b1), the regression results in the original scale: in red, mortality under 1 month (x1); in blue, mortality between 1 month and 5 years (x2); in green, survival beyond 5 years (x3). The shaded areas represent the 90% confidence intervals.

Figure 2: **Mortality estimates for Mauritania, with (a) logit and (b) isometric log-ratio transformations**
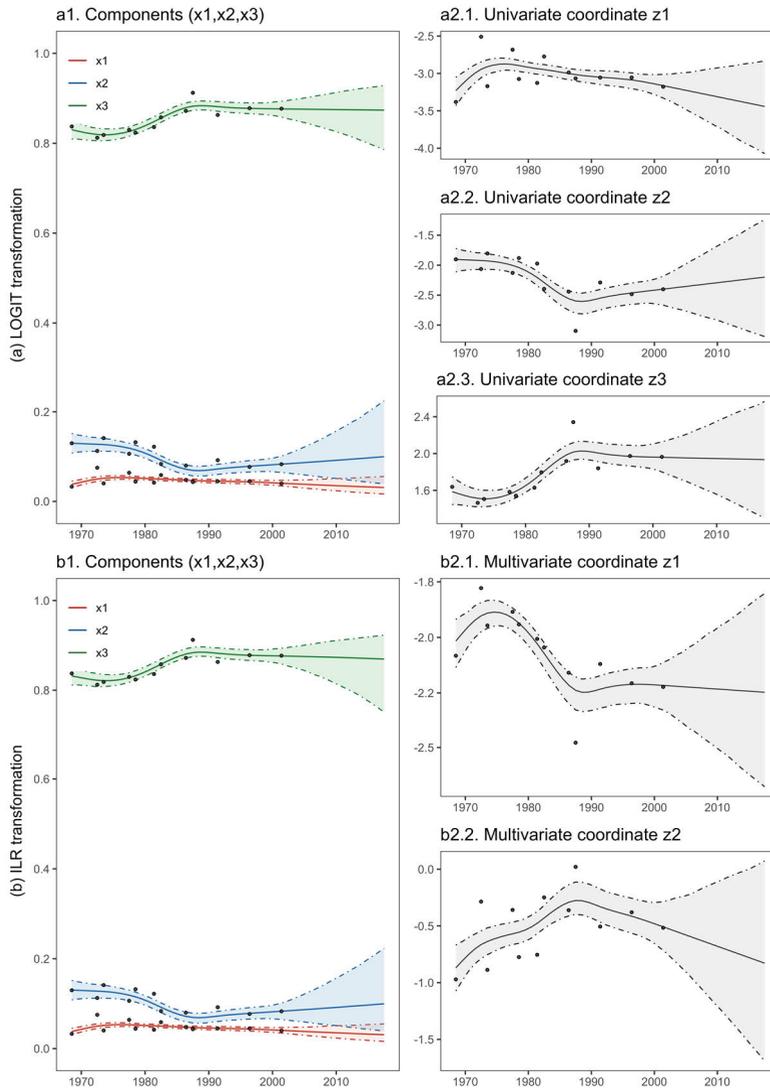
Notes: The regression results in the coordinates are in black: the three univariate coordinates for LOGIT (in a2.1, a2.2 and a2.3) and the two multivariate coordinates for ILR (in b2.1 and b2.2). The regression results in the original scale are in colour (a1 and b1): mortality under 1 month (x_1) in red; mortality between 1 month and 5 years (x_2) in blue; survival beyond 5 years (x_3) in green. The shaded areas represent the 90% confidence intervals.

On the other hand, the ILR data transformation results in two multivariate coordinates: z_1 captures information about the survival rate, while z_2 captures the relationship between the mortalities under 1 month and from ages 1 month to 5 years. Again, the patterns in z_1 and z_2 differ in the two countries. In Malawi, z_1 declines almost steadily from –1.2 in 1970 to –2.7 in 2018, while in z_2 there is no change in the first 23 years but it starts to increase by 0.4 points from 1993 onwards. These values indicate a decline not only in child mortality over the years but specifically in the rates of mortalities under 1 month and between 1 month and 5 years from 1993. In Mauritania, z_1 shows three different phases: from 1968 to 1976 it increases slightly (by 0.1 points), between 1976 and 1991 it falls (by 0.3 points), and from then until 2018 it remains practically unchanged. z_2 first increases by 0.6 points between 1968 and 1989 and then declines until reaching –0.82 in 2018. These results explain the inconspicuous change in child mortality in Mauritania from 1989 and the slight setback in the survival of children aged between 1 month and 5 years.

For both data transformations, the resulting components (i.e., x_1, x_2, and x_3) are substantially the same (Table 3). For instance, in Malawi the estimates for neonatal mortality (x_1) obtained with the LOGIT transformation are 46.5, 37.4, 28.0, and 23.3 per mil for 1990, 2000, 2010, and 2018, respectively; with ILR these estimates are 46.5, 37.5, 27.9, and 22.9. In Mauritania the results are also practically the same as in x_2: 70.8, 82.3, 92.2, and 99.8 with LOGIT versus 70.8, 82.3, 92.2, and 99.7 with ILR.

Table 3: **Mortality estimates, expressed per mil (10^{-3}), for the years 1990, 2000, 2010, and 2018, in Malawi and Mauritania, obtained with LOGIT and ILR transformations**

Country	Approach	Component	1990	2000	2010	2018
Malawi	LOGIT	x_1	46.5 (43.4–49.2)	37.4 (35.0–39.6)	28.0 (23.6–32.5)	23.3 (16.9–31.1)
		x_2	170.4 (161.4–178.2)	119.4 (113.6–124.7)	68.8 (59.9–77.8)	47.9 (36.4–61.2)
		x_3	782.8 (775.3–790.7)	842.4 (837.9–847.5)	903.4 (895.3–911.6)	930.2 (918.5–941)
	ILR	x_1	46.5 (43.4–49.2)	37.5 (35.1–39.7)	27.9 (23.5–32.3)	22.9 (16.6–30.3)
		x_2	170.5 (161.5–178.3)	119.5 (113.8–124.8)	68.7 (59.9–77.7)	47.7 (36.3–61)
		x_3	783.1 (776.0–791.5)	842.9 (838.5–848.3)	903.4 (895.4–911.7)	929.4 (918.1–939.7)
Mauritania	LOGIT	x_1	45.5 (41.6–49.1)	41.2 (35.7–46.7)	34.9 (23.4–51.1)	31.0 (16.8–55.6)
		x_2	70.8 (59–81.1)	82.3 (65.8–97.8)	92.2 (50.4–160.8)	99.8 (39.5–224.3)
		x_3	882.6 (872.7–893.2)	876.6 (861–892.1)	874.9 (823.6–914.3)	873.8 (785.7–928.6)
	ILR	x_1	45.6 (41.8–49.3)	41.2 (35.7–46.7)	34.8 (22.9–51)	30.9 (16.1–54.8)
		x_2	70.8 (58.9–81.1)	82.3 (65.8–97.9)	92.2 (50.4–160.2)	99.7 (39.5–223.1)
		x_3	883.6 (874.1–895.1)	876.5 (861.1–892.4)	873.0 (811–911.5)	869.5 (751.0–922.8)

Note: Component x_1 is the mortality under 1 month; x_2 the mortality between 1 month and 5 years; and x_3 the survival beyond 5 years.

A closer analysis of the differences between LOGIT and ILR estimates for all countries (Figure 3) reveals that the average difference is 0.9 points per mil, which is negligible for all practical purposes. The maximum differences are found in Liberia, reaching 5.28 and 6.03 per mil for the NMR and U5MR, respectively. Furthermore, although somewhat difficult to decipher from Figure 3, the differences in U5MR are slightly higher than in NMR. This is because U5MR is obtained from the sum of x_1 and x_2 estimates, while NMR is directly x_1. Moreover, the differences between ILR and LOGIT are nearly zero for years with available data, since estimates are close to the observed mortality rates.

Figure 3: **Differences, in absolute value, between estimates obtained with logit and isometric log-ratio transformations for (a) neonatal (NMR) and (b) under-5 (U5MR) mortality rates**
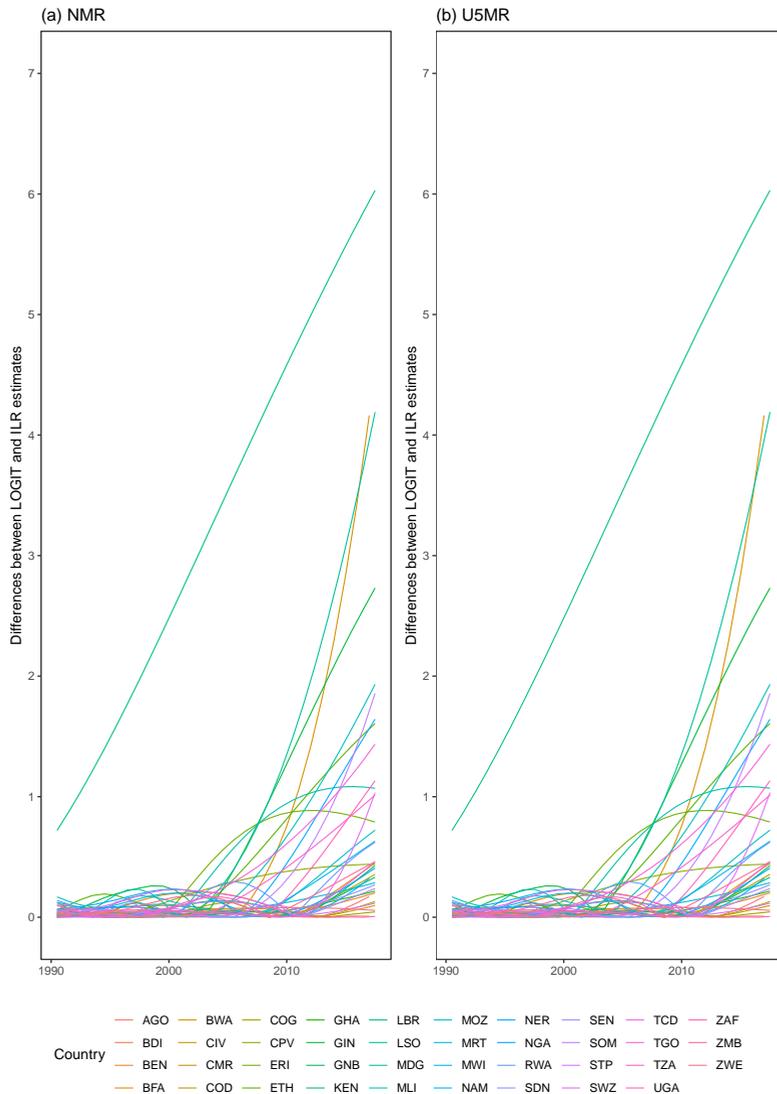
Note: Notice that the countries with the greatest differences (i.e., more than 2 points per mil in 2018) are Liberia (LBR), Botswana (BWA), Lesotho (LSO), and Guinea (GIN).

However, although both LOGIT and ILR transformations yield similar mortality estimates, only the ILR estimates strictly fulfil the unit-sum constraint (Figure 4). In the majority of countries, the sum of mortality estimates with LOGIT is higher than 1. In Lesotho, Botswana, Somalia, and Liberia, for instance, it reaches 1.0951, 1.0584, 1.0305, and 1.0113, respectively (i.e., 95.1, 58.4, 30.5, and 11.3 deaths per 1,000 live births).

This non-unit sum is even more significant for the 90% confidence intervals (Figure 5). For example, in 2018 the 90% confidence intervals of the sum of mortality estimates in these four countries were (1.0015–1.3463) in Lesotho, (1.0060–1.2003) in Botswana, (0.9745–1.5971) in Somalia, and (1.0015–1.1987) in Liberia. By contrast, the sum of the estimates based on ILR-transformed data adds up to 1 in all cases.

Therefore, on theoretical grounds, the LOGIT transformation is not appropriate for child mortality data since it is not consistent with its compositional nature. Indeed, a univariate analysis for each part of the D-part compositions implies merging all remaining parts (i.e., $x_i/1 - x_i$), and such amalgamation does not preserve Aitchison distances in the simplex. As Egozcue and Pawlowsky-Glahn (2005) explain, "distances of amalgamated compositions have a complicated, non-monotonic behaviour with respect to original distances". The change of monotony of distances affects simple operations, such as centring and scaling, and leads to an inconsistent geometric and algebraic representation of compositions. On the contrary, by using $D - 1$ multivariate log-ratios that form an orthonormal basis, the ILR transformation can conserve the metric properties of compositions, and thus their unit-sum constraint. The only case in which LOGIT is valid is when we are dealing with a two-part composition, where the two parts are exclusively associated with each other (i.e., $x_1/x_2 = x_1/(1 - x_1)$ and $x_2/x_1 = x_2/(1 - x_2)$). Indeed, the ILR can be considered a generalization of the LOGIT transformation when compositions have more than two parts (Lloyd, Pawlowsky-Glahn, and Egozcue 2012).

On a practical level, although child mortality estimates are equivalent across the two transformations, this only happens because the data is homogenously balanced: i.e., none of the NMR and U5MR rates approach the boundary values of 0 or 1,000 deaths. As countries continue to reduce their child mortality rates and move closer to 0, the use of the LOGIT transformation will also become more problematic from a practical perspective: child mortality estimates can be misleading at best, and uninterpretable at worst.

Figure 4: **Sum of mortality estimates with (a) logit and (b) isometric log-ratio transformations**

Figure 5: **Sum of mortality estimates with logit and isometric log-ratio transformations in: (a) Lesotho, (b) Botswana, (c) Somalia, and (d) Liberia**
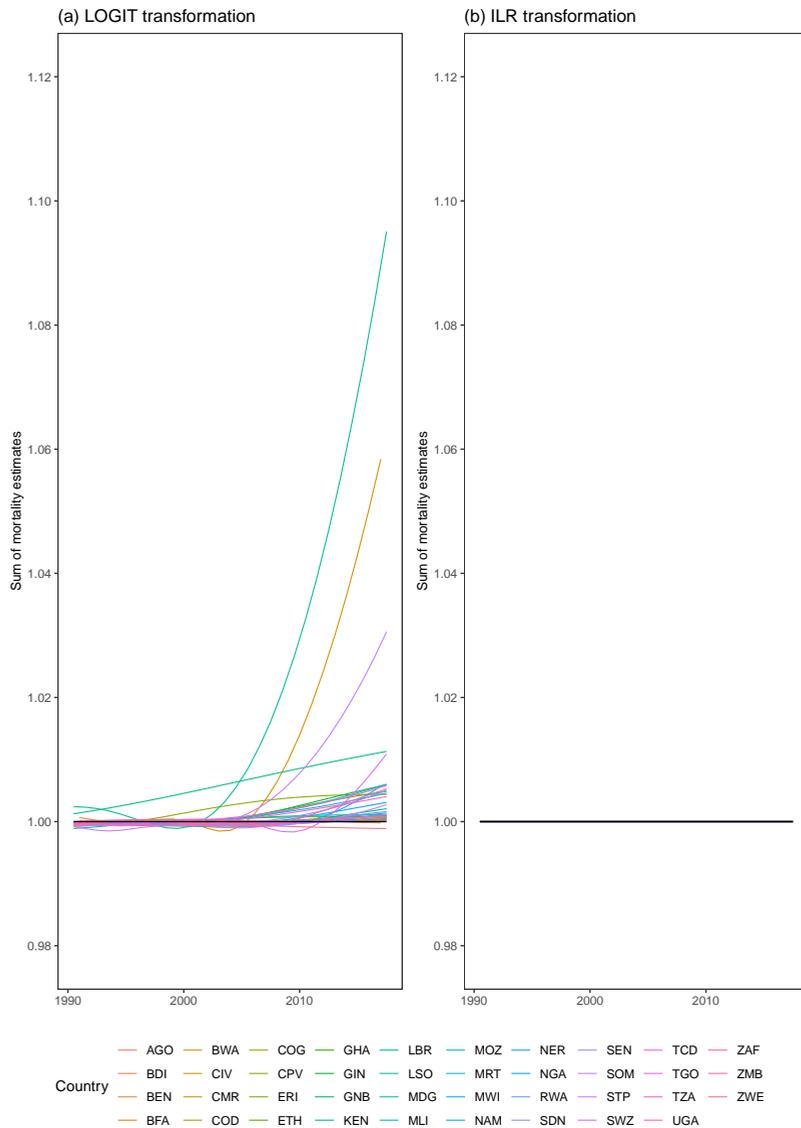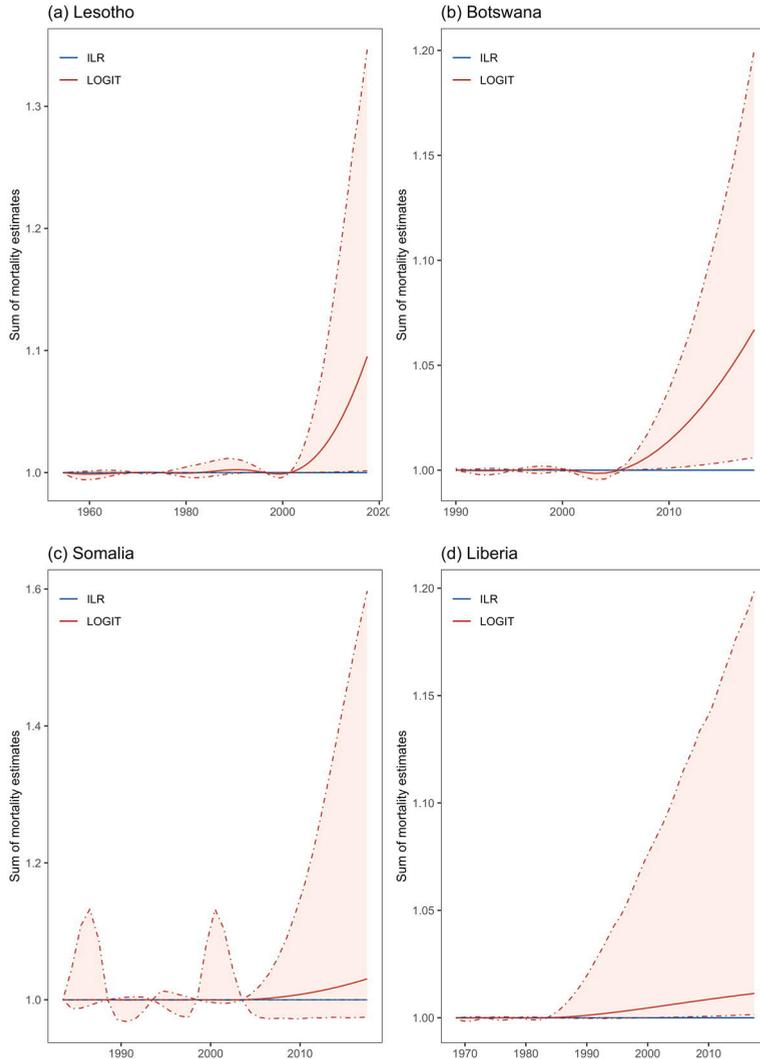
Notes: The LOGIT estimates are in red; the ILR estimates are in blue. The shaded areas represent the 90% confidence intervals. Only the ILR transformation (in blue) fulfils the unit-sum constraint for both observed and simulated data. In years with data the LOGIT transformation (in red) provides estimates closer to the unit sum (hence the narrower confidence intervals).

### 4.2 Comparative analysis 2: ILR vs. UN IGME estimates

Figure 6 shows the comparison of the NMR and U5MR obtained with ILR and those provided by UN IGME for Malawi, Mauritania, Lesotho, and Liberia. As expected, estimates differ significantly between the two approaches, especially regarding the trends and their confidence intervals.

However, the comparison between ILR and UN IGME estimates is not direct, for three main reasons. First, the UN IGME employs a Bayesian penalized spline regression with B-splines, in which information on spline coefficients is exchanged across countries and time periods. This information exchange allows for a better assessment of child mortality trends in countries or periods with limited data. We, on the other hand, fit a simple generalized additive model (GAM) to each country separately. Second, unlike us, the UN IGME's B3 model adjusts for biases due to HIV/AIDS-related mortality,[8] which can confound trend analysis.

Third and foremost, ILR and UN IGME estimates rely on different input data. As detailed in Section 3.2, our regressions only consider year series with observations for both NMR and U5MR (i.e., those that for a given year include information on both mortality rates). In the case of Malawi, for example, our estimates are based on 25 data points, whereas the UN IGME regression model uses all 164 and 35 observations available for the estimation of NMR and U5MR, respectively. This mismatch in the input data becomes more evident in countries like Lesotho and Liberia, where ILR estimates are constructed with merely 5 and 4 observations for each mortality rate, resulting in utterly different NMR and U5MR trends. For instance, in Lesotho, ILR estimates show a substantial increase in U5MR from 2000 onwards (by nearly 170 deaths per 1,000 live births), mainly because the latest observations are excluded from the analysis. These fewer data also lead to wider confidence intervals, especially in later years. In the case of Liberia, our estimates in 2018 are 40.4 for NMR and 84.8 for U5MR, with 90% confidence intervals of (5.5–192.4) and (19.7–534.2). This translates into interval widths of 187.0 and 514.5 deaths per 1,000 live births for NMR and U5MR, respectively. On the contrary, the widths of the UN IGME's confidence intervals are only 25.3 and 51.3 deaths per 1,000 live births for NMR and U5MR, respectively.

---

[8] In populations severely affected by HIV/AIDS (i.e., those where the prevalence reaches 5% of the adult population), there is a correlation between the mortality risk of mothers and their children: HIV-positive children are more likely to die than other children, and are less likely to be reported, since their mothers are more likely to die also. Therefore, child mortality estimates are biased downwards. That is why the UN IGME adjusts for bias due to AIDS in child mortality estimation.

Figure 6: **Neonatal (NMR) and under-5 (U5MR) mortality rates for:
(a) Malawi, (b) Mauritania, (c) Lesotho and (d) Liberia**
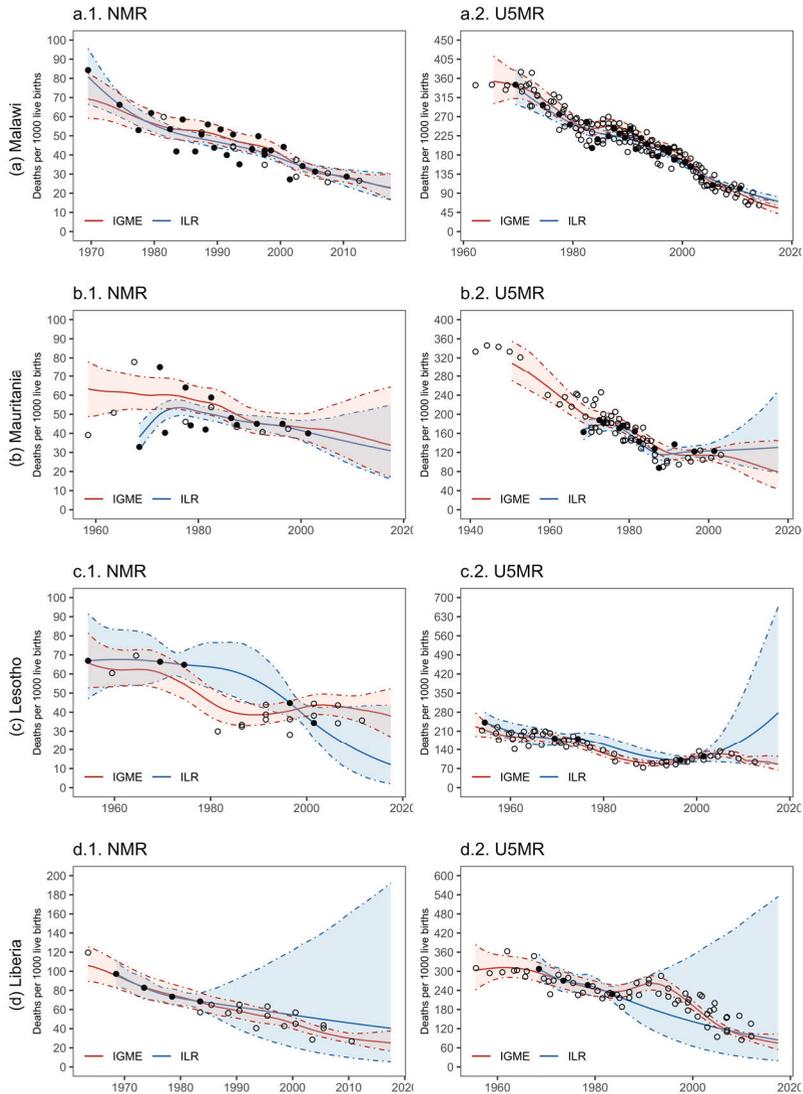
Notes: Estimates obtained with the isometric log-ratio transformation are in blue; those provided by the UN IGME are in red. The shaded areas represent the 90% confidence intervals. The solid dots represent the data points used by both ILR and the UN IGME's model, while the hollow dots represent those used only by the UN IGME's model.

The UN IGME's B3 model is thus more sophisticated and complete than the one we employ. However, it suffers from one key drawback: it employs LOG and LOGIT transformations to child mortality rates, which, as demonstrated in the previous section (4.1), do not guarantee consistent distances and statistical analysis when working with compositional data. This inconsistency might go unnoticed in the current child mortality estimates provided by the UN IGME, where survival rates beyond 5 years are not reported and hence the unit-sum constraint cannot be checked. Yet this remains an important issue to be addressed, especially from a theoretical standpoint. Moreover, as countries continue to witness a decline in child mortality and estimates move towards values closer to zero, the issue of compositional incoherence can further exacerbate spurious correlations and misleading inferences.

A compositional approach circumvents this pitfall, but at the expense of data availability. As we have seen, to undertake a compositional analysis of child mortality we must exclude time series with incomplete data from the regression. This has a substantial impact on the number of observations considered, since merely 18% of the time-series from the UN IGME database include observations for both NMR and U5MR in sub-Saharan Africa. The reason why there is less data available for NMR than for U5MR is simple: neonatal mortality cannot be indirectly estimated from summary birth histories, as is done for under-5 child mortality (Burstein et al. 2018). In extreme cases such as Gambia, where there is no complete time series, the country is omitted altogether from the analysis. In other cases, the limited data available for NMR means that regressions are based on very few observations. Thus, ILR-based trends should be taken with caution, especially when comparing them to those provided by the UN IGME.

Given the scarcity of complete data for most developing countries, the critical question is therefore whether we should prioritize more conceptually sound mortality estimates, even if this entails significantly fewer observations in the regression analysis and hence less interpretable trends. The answer to this question is not that straightforward. On the one hand, theoretically inconsistent estimates can result in a misleading overview of the health situation and child mortality trends, especially when the data are close to their range limits (0 or 1,000 deaths). On the other hand, trend analysis based on few observations must be interpreted with care, as the risks of getting nonsensical results are relatively high.

The ideal solution to this intricate problem lies in increasing the amount of complete data available for both NMR and U5MR. However, due to the prohibitive and time-consuming nature of household surveys, the most cost-effective option requires better utilization of existing data. This may be done in two ways. First, missing values of NMR or U5MR can be filled by imputation techniques (both parametric and nonparametric), but using replacement strategies that do not distort the covariance structure of the involved parts (Martín-Fernández, Barceló-Vidal, and Pawlowsky-Glahn 2003; Quispe-

Coica and Pérez-Foguet 2020). Second, data availability for neonatal mortality – which is the indicator with the fewest observations – can be improved by developing and validating new indirect estimation methods that derive rates from summary birth histories, as is currently done for under-5 mortality. As Burstein et al. (2018) highlight, "[the] use of such methods allows research to utilize a massive amount of SBH data for the estimation of trends in neonatal mortality" and consequently "further improve the evidence base for monitoring of trends and inequalities".

4.3 Geographical analysis: Patterns in sub-Saharan Africa

The 1990–2018 evolution of the neonatal (NMR) and under-5 (U5MR) mortality rates of sub-Saharan African countries based on our ILR compositional approach are shown in Figure 7. Child mortality rates show a substantial decline in the last 30 years in all sub-Saharan African countries except for Botswana, Somalia, and Sudan in the case of NMR, and Lesotho, and Togo in the case of U5MR. However, much of this rise in child mortality is due to the very few data points used in the regression. In Botswana, for instance, only 4 data points are included in the model, the most recent being in 2005. Furthermore, in nearly half of the countries, the analysis is done with less than 10 observations (as seen in Table 2), which hinders the reliability of mortality estimates. Consequently, in countries with limited data, patterns in child mortality should be taken with caution.

On the other hand, in 2018 only three countries – Eritrea, Cape Verde, and Madagascar (7.9% of all countries) – had already met the SDG 3 target 3.2 of reducing neonatal and under-5 mortality rates to 12 and 25 deaths per 1,000 lives births by 2030, respectively. The distribution of NMR for the other countries is as follows: 44.7% of them are between the target value and two-times the target value (i.e., between 12 and 24 deaths), 34.2% are between two- and three-times the target value (24 and 36 deaths), and 13.2% had rates over triple the target. For U5MR, these figures are 23.7%, 26.3%, and 42.1%, respectively, which presents a far less hopeful picture of under-5 mortality.

Furthermore, a geographical analysis of child mortality estimates shows notable disparities between sub-Saharan African regions. Maps of both the NMR and U5MR rates show a concentration of high mortality in the regions of Western and Central Africa. For instance, in 2018 the average NMR in countries of Western and Central Africa is higher than 25 per million, while the regional average in Southern and Eastern Africa (excluding Botswana and Somalia) is less than 18.

Figure 7: **Evolution of child mortality in sub-Saharan Africa in 1990, 2000, 2010, and 2018, for (a) neonatal (NMR) and (b) under-5 (U5MR) mortality rates, based on the estimates obtained with isometric-log ratio transformation**

(a) NMR (Deaths per 1,000 live births)   (b) U5MR (Deaths per 1,000 live births)
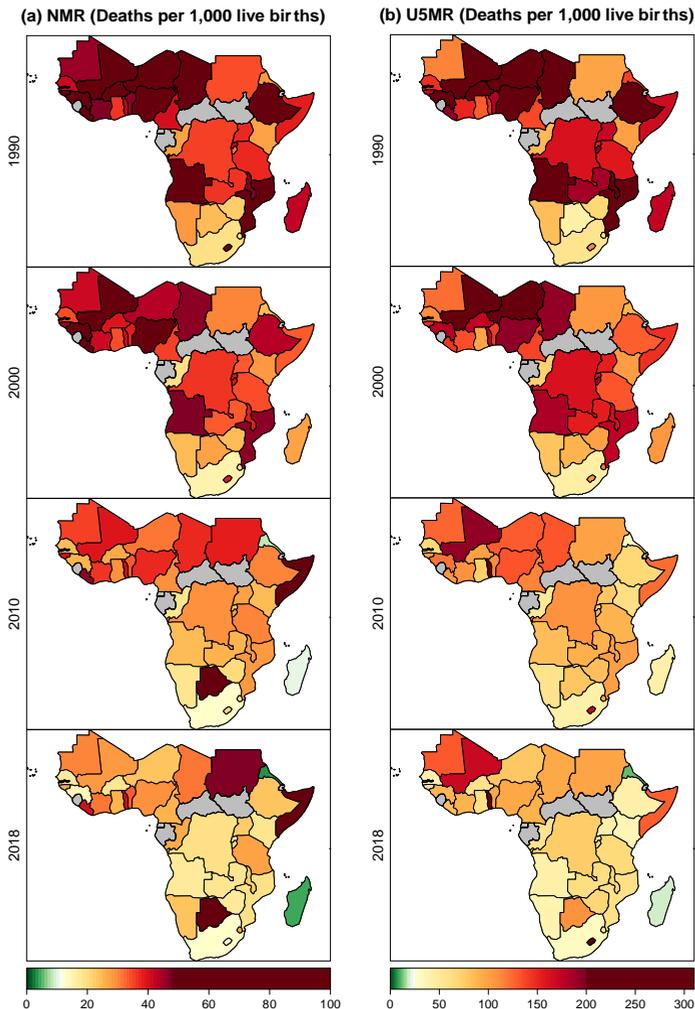
Notes: Countries with mortality rates lower than the SDG 3 targets are in green. Countries with mortality rates greater than the target values are in red. Countries with insufficient data that are excluded from the analysis are In grey, Notice that the colour ranges are different for the NMR and U5MR maps because of the different SDG 3 targets (i.e., 12 and 25 deaths per 1,000 live births for NMR and U5MR, respectively).

One immediate policy implication can be drawn from this geographical analysis: to achieve the SDG 3 targets for child mortality, policymakers should dedicate more means to increasing access to and use of maternal and childcare services, especially in Western and Central African countries. In turn, this requires a stronger political commitment from the national governments along with international support.

5. Conclusion

The estimation of child mortality is challenging for the vast majority of developing countries, where vital registration systems are often incomplete or unreliable and thus models are required to construct neonatal (NMR) and under-5 (U5MR) mortality estimates for years of interest. Such child mortality rates are, by definition, compositional: the individual rates of children dying in the different age intervals (i.e., 0 to 28 days and 29 days to 5 years) and surviving the age of 5 are not independent of each other but related by being probabilities.

In this paper, we explore the application of compositional data analysis to child mortality estimation. In particular, we compare two data transformation approaches – the log of the odds (LOGIT) and the isometric log-ratio (ILR) – and assess their theoretical and practical suitability for estimating child mortality levels and trends.

Three key findings emerge from our study. First, while the LOGIT transformation is widely used in child mortality estimation (for instance, in the UN IGME's B3 model), it is only theoretically valid to analyze two-part compositions. When applied to three-part compositions such as child mortality data, LOGIT does not address the compositional characteristics inherent in probabilities, such as the unit-sum constraint. Second, the ILR transformation leads to more conceptually sound results, but to the detriment of data availability. A compositional approach requires the exclusion of incomplete time series with missing values for NMR or U5MR, which is relatively common in child mortality data. This loss of data limits the ability to produce plausible trends for some countries. Third, two solutions are possible to fill this critical gap in data on child mortality and improve its trend analysis. First, missing values in the current UN IGME dataset can be replaced through imputation techniques that preserve the compositional structure of the data. Second, data for NMR – which is the indicator with fewer available observations – can be improved with new indirect estimation methods that expand the potential utility of summary birth history data.

However, it is worth noting that our proposal does not propose replacing the UN IGME's model with our ILR-based regression approach, but rather underscores the need to account for the compositional nature of child mortality data. It is clear that our model is less sophisticated than the well-established B3 model, but it contains some new features

that can be incorporated for more theoretically robust modelling of child mortality. In this sense, future research should explore new methods for Bayesian penalized regression modelling of compositional data.

6. Acknowledgements

We thank the two anonymous reviewers for their critical evaluation and insightful feedback throughout the review processes.

6.1 Author contributions

Conceptualization: F.E., A.P.F. Methodology: F.E., A.P.F. Software: F.E. Data analysis and visualization: F.E. Writing – original draft: F.E. Writing – reviewing and editing: F.E., A.P.F. Supervision: A.P.F.

6.2 Funding

References

Aitchison, J. (1982). The statistical analysis of compositional data. *Journal of the Royal Statistical Society: Series B (Methodological)* 44(2): 139–160. https://www.jstor.org/stable/2345821.

Aitchison, J. (1999). Logratios and natural laws in compositional data analysis. *Mathematical Geology* 31(5): 563–580. doi:10.1023/A:1007568008032.

Alexander, M. and Alkema, L. (2018). Global estimation of neonatal mortality using a Bayesian hierarchical splines regression model. *Demographic Research* 38(15) 335–372. doi:10.4054/DemRes.2018.38.15.

Alkema, L. and New, J.R. (2014). Global estimation of child mortality using a Bayesian B-spline bias-reduction model. *Annals of Applied Statistics* 8(4): 2122–2149. doi:10.1214/14-AOAS768.

Alkema, L., Chao, F., You, D., Pedersen, J., and Sawyer, C.C. (2014b). National, regional, and global sex ratios of infant, child, and under-5 mortality and identification of countries with outlying ratios: a systematic assessment. T*he Lancet Global Health* 2(9): e521–e530. doi:10.1016/S2214-109X(14)70280-3.

Alkema, L., New, J.R., Pedersen, J., and You, D. (2014a). Child mortality estimation 2013: An overview of updates in estimation methods by the United Nations Inter-agency Group for Child Mortality Estimation. *PloS One* 9(7): e101112. doi:10.1371/journal.pone.0101112.

Ayele, D.G., Zewotir, T.T., and Mwambi, H.G. (2015). Structured additive regression models with spatial correlation to estimate under-five mortality risk factors in Ethiopia. *BMC Public Health* 15(1): 268. doi:10.1186/s12889-015-1602-z.

Bergeron-Boucher, M.P., Canudas-Romo, V., Oeppen, J., and Vaupel, J.W. (2017). Coherent forecasts of mortality with compositional data analysis. *Demographic Research* 37(17): 527–566. doi:10.4054/DemRes.2017.37.17.

Bergeron-Boucher, M.P., Simonacci, V., Oeppen, J., and Gallo, M. (2018). Coherent modeling and forecasting of mortality patterns for subpopulations using multiway analysis of compositions: An application to Canadian provinces and territories. *North American Actuarial Journal* 22(1): 92–118. doi:10.1080/10920277.2017.1377620.

Bermejo, R., Firth, S., Hodge, A., Jimenez-Soto, E., and Zeck, W. (2015). Overcoming stagnation in the levels and distribution of child mortality: The case of the Philippines. *PloS One* 10(10): e0139458. doi:10.1371/journal.pone.0139458.

Brass, W. (1971). Methods for estimating fertility and mortality from limited and defective data. Technical report. Chapel Hill: The University of North Carolina at Chapel Hill, International Program for Laboratories for Population Statistics.

Burstein, R., Wang, H., Reiner Jr, R.C., and Hay, S.I. (2018). Development and validation of a new method for indirect estimation of neonatal, infant, and child mortality trends using summary birth histories. *PLoS Medicine* 15(10): e1002687. doi:10.1371/journal.pmed.1002687.

Carson, V., Tremblay, M.S., Chaput, J.P., and Chastin, S.F. (2016). Associations between sleep duration, sedentary time, physical activity, and health indicators among Canadian children and youth using compositional analyses. *Applied Physiology, Nutrition, and Metabolism* 41(6): S294–S302. http://doi.org/10.1139/apnm-2016-0026.

Chao, F., You, D., Pedersen, J., Hug, L., and Alkema, L. (2018). National and regional under-5 mortality rate by economic status for low-income and middle-income countries: A systematic assessment. *The Lancet Global Health* 6(5): e535–e547. doi:10.1016/S2214-109X(18)30059-7.

Dlamini, W.J., Melesse, S.F. and Mwambi, H.G. (2017). Logistic regression additive model: Application to Tanzania demographic and health survey data. *The Open Public Health Journal* 10(1). doi:10.2174/1874944501710010294.

Dwyer-Lindgren, L., Kakungu, F., Hangoma, P., Ng, M., Wang, H., Flaxman, A.D., Masiye, F., and Gakidou, E. (2014). Estimation of district-level under-5 mortality in Zambia using birth history data, 1980–2010. *Spatial and Spatio-Temporal Epidemiology* 11: 89–107. doi:10.1016/j.sste.2014.09.002.

Egozcue, J.J. and Pawlowsky-Glahn, V. (2005). Groups of parts and their balances in compositional data analysis. *Mathematical Geology* 37(7): 795–828. doi:10.1007/s11004-005-7381-9.

Egozcue, J.J., Pawlowsky-Glahn, V., Mateu-Figueras, G., and Barcelo-Vidal, C. (2003). Isometric logratio transformations for compositional data analysis. *Mathematical Geology* 35(3): 279–300. doi:10.1023/A:1023818214614.

Ezbakhe, F. and Pérez-Foguet, A. (2019). Compositional analysis of child mortality (R code). Available at: doi:10.5281/zenodo.4027068.

Ezbakhe, F. and Pérez-Foguet, A. (2019). Estimating access to drinking water and sanitation: The need to account for uncertainty in trend analysis. *Science of the Total Environment* 696. doi:10.1016/j.scitotenv.2019.133830.

Fairclough, S.J., Dumuid, D., Taylor, S., Curry, W., McGrane, B., Stratton, G., Maher, C., and Olds, T. (2017). Fitness, fatness and the reallocation of time between children's daily movement behaviours: An analysis of compositional data. *International Journal of Behavioral Nutrition and Physical Activity* 14(1): 64. doi:10.1186/s12966-017-0521-z.

Golding, N., Burstein, R., Longbottom, J., Browne, A.J., Fullman, N., Osgood-Zimmerman, A., Earl, L., Bhatt, S., Cameron, E., Casey, D.C., and Dwyer-Lindgren, L. (2017). Mapping under-5 and neonatal mortality in Africa, 2000–15: A baseline analysis for the Sustainable Development Goals. *The Lancet* 390(10108): 2171–2182. doi:10.1016/S0140-6736(17)31758-0.

Guillot, M., Gerland, P., Pelletier, F., and Saabneh, A. (2012). Child mortality estimation: A global overview of infant and child mortality age patterns in light of new empirical data. *PLoS Medicine* 9(8): e1001299. doi:10.1371/journal. pmed.1001299.

Hastie, T. (2018). Gam: Generalized additive models. R package version 1.16. Available at: https://cran.r-project.org/web/packages/gam/.

Hill, K., Brady, E., Zimmerman, L., Montana, L., Silva, R., and Amouzou, A. (2015). Monitoring change in child mortality through household surveys. *PloS One* 10(11): e0137713. doi:10.1371/journal.pone.0137713.

Hill, K., You, D., Inoue, M., and Oestergaard, M.Z. (2012). Child mortality estimation: Accelerated progress in reducing global child mortality, 1990–2010. *PLoS Medicine* 9(8): e1001303. doi:10.1371/journal.pmed.1001303.

Hodge, A., Firth, S., Marthias, T., and Jimenez-Soto, E. (2014). Location matters: trends in inequalities in child mortality in Indonesia: Evidence from repeated cross-sectional surveys. *PloS One* 9(7): e103597. doi:10.1371/journal.pone.0103597.

Kirigia, J.M., Muthuri, R.D.K., Nabyonga-Orem, J., and Kirigia, D.G. (2015). Counting the cost of child mortality in the World Health Organization African region. *BMC Public Health* 15(1): 1103. doi:10.1186/s12889-015-2465-z.

Korenromp, E.L., Arnold, F., Williams, B.G., Nahlen, B.L., and Snow, R.W. (2004). Monitoring trends in under-5 mortality rates through national birth history surveys. *International Journal of Epidemiology* 33(6): 1293–1301. doi:10.1093/ ije/dyh182.

Lesser, J. and Kalsbeek, W. (1999). Non-sampling error in environmental surveys. *Journal of Agricultural, Biological, and Environmental Statistics* 4(4): 473–488. https://www.jstor.org/stable/1400503.

Li, Z., Hsiao, Y., Godwin, J., Martin, B.D., Wakefield, J., and Clark, S.J. (2019). Changes in the spatial distribution of the under-five mortality rate: Small-area analysis of 122 DHS surveys in 262 subregions of 35 countries in Africa. *PloS One* 14(1): e0210645. doi:10.1371/journal.pone.0210645.

Lloyd, C.D., Pawlowsky-Glahn, V., and Egozcue, J.J. (2012). Compositional data analysis in population studies. *Annals of the Association of American Geographers* 102(6): 1251–1266. doi:10.1080/00045608.2011.652855.

Mahapatra, P., Shibuya, K., Lopez, A.D., Coullare, F., Notzon, F.C., Rao, C., and Szreter, S. (2007). Civil registration systems and vital statistics: Successes and missed opportunities. *The Lancet* 370(9599): 1653–1663. doi:10.1016/S0140-6736(07)61308-7.

Marcillo-Delgado, J.C., Ortego, M.I., and Pérez-Foguet, A. (2019). A compositional approach for modelling SDG7 indicators: Case study applied to electricity access. *Renewable and Sustainable Energy Reviews* 107: 388–398. doi:10.1016/j.rser.2019.03.028.

Martín-Fernández, J.A., Barceló-Vidal, C., and Pawlowsky-Glahn, V. (2003). Dealing with zeros and missing values in compositional data sets using nonparametric imputation. *Mathematical Geology* 35(3): 253–278. doi:10.1023/A:1023866030544.

Mercer, L., Wakefield, J., Pantazis, A., Lutambi, A., Masanja, H., and Clark, S. (2015). Small area estimation of child mortality in the absence of vital registration. *The Annals of Applied Statistics* 9(4): 1889–1905. doi:10.1214/15-AOAS872.

Minnery, M., Firth, S., Hodge, A., and Jimenez-Soto, E. (2015). Neonatal mortality and inequalities in Bangladesh: Differential progress and sub-national developments. *Maternal and Child Health Journal* 19(9): 2038–2047. doi:10.1007/s10995-015-1716-z.

Oeppen, J. (2008). *Coherent forecasting of multiple-decrement life tables: A test using Japanese cause of death data*. Paper presented at the European Population Conference 2008, Barcelona, Spain, July 9–12, 2008. http://hdl.handle.net/10256/742.

Pawlowsky-Glahn, V. and Egozcue, J.J. (2006). Compositional data and their analysis: An introduction. *Geological Society, London, Special Publications* 264(1): 1–10. doi:10.1144/GSL.SP.2006.264.01.01.

Pawlowsky-Glahn, V., Egozcue, J.J., and Tolosana-Delgado, R. (2015). *Modeling and analysis of compositional data*. London: John Wiley & Sons.

Pérez-Foguet, A., Giné-Garriga, R. and Ortego, M.I. (2017). Compositional data for global monitoring: The case of drinking water and sanitation. *Science of the Total Environment* 590: 554–565. doi:10.1016/j.scitotenv.2017.02.220.

Pezzulo, C., Bird, T., Utazi, E.C., Sorichetta, A., Tatem, A.J., Yourkavitch, J., and Burgert-Brucker, C.R. (2016). Geospatial modeling of child mortality across 27 countries in sub-Saharan Africa. DHS Spatial Analysis Reports 13. USAID. Available at: https://eprints.soton.ac.uk/400944/1/SAR13.pdf.

Quispe-Coica, A. and Pérez-Foguet, A. (2020). Preprocessing alternatives for compositional data related to water, sanitation and hygiene. *Science of The Total Environment* 743: 140519. doi:10.1016/j.scitotenv.2020.140519.

Rajaratnam, J.K., Tran, L.N., Lopez, A.D., and Murray, C.J. (2010). Measuring under-five mortality: Validation of new low-cost methods. *PLoS Medicine* 7(4): e1000253. doi:10.1371/journal.pmed.1000253.

Rutstein, S.O. (1984). Infant and child mortality: Levels trends and demographic differentials: World Fertility Survey (WFS). *Comparative Studies* 43. Voorburg: International Statistical Institute.

Rutstein, S.O. and Rojas, G. (2006). Guide to DHS statistics. Technical report. Calverton, MD: Macro ORC. Available at: http://www.ophi.org.uk/wp-content/uploads/OPHI-HDCA-SS2012-Guide_DHS_Statistics.pdf.

Salomon, J.A. and Murray, C.J.L. (2001) Compositional models for mortality by age, sex and cause. Geneva: World Health Organization (GPE Discussion Paper Series 11) Available at: https://www.who.int/healthinfo/paper11.pdf.

UN IGME (Inter-agency Group for Child Mortality Estimation) (2018). Explanatory Notes: Child mortality trend series to 2017. New York: United Nations. Available at: https://childmortality.org/wp-content/uploads/2018/12/English_IGME_country_consultation_note.pdf.

UNESA (United Nations Department of Economic and Social Affairs) (2019). Population Division: World population prospects 2019. New York: United Nations. Available at: https://population.un.org/wpp/DataQuery/.

UNESA (United Nations Department of Economic and Social Affairs) (2011). Mortality estimates from major sample surveys: towards the design of a database for the monitoring of mortality levels and trends. Technical Paper 2011/2. New York: United Nations. Available at: https://www.un.org/en/development/desa/population/publications/pdf/technical/TP2011-2_MortEstMajorSampSurv.pdf.

UNGA (United Nations General Assembly) (2015). Transforming our world: The 2030 Agenda for Sustainable Development. A/RES/70/1. New York: United Nations. Available at: https://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E.

UNICEF (United Nations Children's Fund) (2006). Multiple Indicator Cluster Survey Manual 2005: Monitoring the Situation of children and women. UNICEF's Division of Policy and Planning. New York: United Nations. Available at: http://mics.unicef.org/files?job=W1siZiIsIjIwMTUvMDQvMDIvMDYvMzcvMDYvMTE5L011bHRpcGxlX0luZGljYXRvcl9DbHVzdGVyX1N1cnZleV9NYW51YWxfMjAwNS5wZGYiXV0&sha=dd2e54d1ddd61cdb.

UNICEF (United Nations Children's Fund) (2017). Early experiences with the Equitable Impact Sensitive Tool – EQUIST. Health System Strengthening Unit Working Paper. New York: UNICEF. Available at: https://www.equist.info/files/general_files/GLB_7834equist-working-paper-jan-2018.pdf.

UNICEF (United Nations Children's Fund) (2019). Levels and trends in child mortality report 2019. New York: UNICEF. Available at: https://data.unicef.org/resources/levels-and-trends-in-child-mortality.

van den Boogaart, G.K., Tolosana-Delgado, R., and Bren, M. (2018). Compositions: Compositional data analysis. R package version 1.40-2. Available at: https://cran.r-project.org/web/packages/compositions.

Wakefield, J., Fuglstad, G.A., Riebler, A., Godwin, J., Wilson, K., and Clark, S.J. (2018). Estimating under-five mortality in space and time in a developing world context. *Statistical Methods in Medical Research* 28(9): 2614–2634. doi:10.1177/0962280218767988.

Wang, H., Liddell, C.A., Coates, M.M., Mooney, M.D., Levitz, C.E., Schumacher, A.E., Apfel, H., Iannarone, M., Phillips, B., Lofgren, K.T., and Sandar, L. (2014). Global, regional, and national levels of neonatal, infant, and under-5 mortality during 1990–2013: A systematic analysis for the Global Burden of Disease Study 2013. *The Lancet* 384(9947): 957–979. doi:10.1016/S0140-6736(14)60497-9.

Wood, S.N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 65(1): 95–114. doi:10.1111/1467-9868.00374.

Zlotnik, H. and Hill, K. (1981). The use of hypothetical cohorts in estimating demographic parameters under conditions of changing fertility and mortality. *Demography* 18(1): 103–122. doi:10.2307/2061052.