Aging of measurement in the association between anthropometry and mortality in children

David Roodman

Senior Advisor, Open Philanthropy & GiveWell

December 30, 2022

Abstract

Background. Millions of children under 60 months old worldwide are undernourished. Olofin et al. (2013) investigates the risks associated with suboptimal growth in such children using older data from prospective cohort studies in 10 countries. The Cox proportional hazards model is fit to each data set, and the results pooled meta-analytically.

Methods. To engage critically with Olofin, the study is replicated. 5 of the 10 data sets are obtained, as well as one not used in Olofin. The original results are approximately replicated. A methodological issue is exposed: the Cox model's assumption that hazard ratios are fixed over time is violated by aging of measurement. Hazard ratios are higher in data sets with more frequent follow-up. Two alternative models are applied: a probit model for death within one year of measurement, and a stochastic process ("MIG") model.

Results. The two new methods yield one-year mortality RRs that are lower than some corresponding HRs in Olofin. Where Olofin reports an HR for weight-for-height *z* (WHZ) < –3 vs. WHZ > –1 of 11.63, the probit and MIG RRs for WHZ = –3.5 vs. –0.5 are 3.75 (2.22, 6.31) and 5.61 (2.58, 12.18).

Conclusions. For coherence, pooled analyses of historical data with various follow-up frequencies should employ methods that are less sensitive to timing effects and should state results with reference to specified follow-up periods. HRs and RRs are not constant.

Introduction

As an input to GiveWell's assessment of community-based management of acute malnutrition (CMAM), I reviewed and reanalyzed literature on the statistical association between anthropometric indicators of nutritional status in children and subsequent mortality. The starting point was Olofin et al. (2013). That review gathers data from previous studies in ten countries, fits Cox proportional hazards models to each, combines the results with random-effects meta-analysis, and reports hazard ratios for various anthropometric groups. Many of the studies from which data are taken focus on other questions, such as the benefit of Vitamin A supplementation. All repeatedly measure young children in community settings before the advent of CMAM, between the mid-1970s and mid-1990s.

The present review works with five of the Olofin data sets, as well as one from the Democratic Republic of Congo (Van den Broeck, Eeckels, and Vuylsteke 1993; Schwinger et al. 2019).[1] The main finding is that *aging of measurement is likely a major and unrecognized source of disagreement across data sets on the size of the relationship between anthropometry and death*. When standard methods are applied to the data set with the most frequent follow-up—Adair et al. (1993), from the Philippines, with bimonthly measurements

[1] A note on the ethics of this work: the Harvard School of Public Health's Institutional Review Board assessed Olofin's pooled analysis of anonymized human subjects data as exempt from ethics requirements (Olofin et al. 2013, p. 2).

of children—the methods are assessing the risks associated with anthropometry taken 0–2 months before, or 1 month ago on average. When they are applied to the data from Senegal (Garenne et al. 2000), most of which was gathered at 6 months intervals, the methods are assessing a different set of risks, those associated with anthropometry taken on average at least 3 months ago. Just as a Covid test result quickly loses relevance, the hazard ratios associated with indicators of wasting decline with time since measurement. This probably explains why in the Olofin et al. (2013, Table S2(C)) review, the Philippines data produces the highest hazard ratio for weight-for-height *z* score (WHZ) < –3 and the Senegal data the lowest: 39.00 vs. 5.41. (According to a World Health Organization standard, WHZ < –3 justifies a diagnosis of severe acute malnutrition, or SAM.)

To improve comparability of risk and hazard ratios from data sets with differing follow-up frequencies, this paper develops and applies two philosophically opposite methods, which turn out to generate similar results. The first approach throws away most of the timing information in the data sets in order to homogenize them. In particular, where data permit, single, one-year spells are constructed for each child. Each begins with the child's first measurement after age 6 months. Probit models are fit to the binary survival outcome; the timing of deaths within that year plays no role.

The other approach retains all the timing information and marshals a more sophisticated estimation model that, to the extent that its mathematical assumptions hold, is robust to changes in follow-up frequency. This is the stochastic process survival model of Aalen (1994). I call it the MIG model because it generates a mixture inverse Gaussian distribution for time of failure. The model generates unobserved heterogeneity by endowing subjects with "health trajectories" with random slopes within each episode. Hazards and hazard ratios fluctuate in time since last measurement, as subjects with downward slopes experience a burst of mortality early on and the survivors then escape most risk of death.

For the sake of consistent interpretation, and with an eye toward assessing the impact of CMAM programs, I focus on ratios in cumulative mortality in the year following measurement that are predicted by fitted models. In particular, I compare WHZ = –3.5, a representative value for WHZ-indicated SAM (WHZ-SAM), to a far healthier –0.5. Similarly, using the other recognized diagnostic for SAM, mid-upper arm circumference (MUAC), I compare MUAC = 11 to MUAC = 14. These ratios correspond to the most eye-catching numbers in Olofin et al. (2013) and similar studies. In addition, I make two narrower comparisons that may be more relevant for the operation of CMAM. In Demographic and Health Surveys (DHS) data for eight African nations, WHZ among children with SAM or moderate acute malnutrition averages –2.77.[2] And (preliminarily) it appears that combined CMAM programs, which serve both MAM and SAM children, typically raise WHZ by a point. I therefore also report risk ratios for WHZ = –2.75 vs. –1.75 and, correspondingly, MUAC = 11.75 vs. 12.75. The meta-analytical results are gathered in Table 1. Predicted mortality ratios are statistically consistent across method and metric: about 3.5–5.5 for the wider comparison and 1.6–1.8 for the narrower one.

It is possible that the causal impact of malnourishment on mortality in these settings was much smaller or larger than these associations suggest. That possibility is to a substantial degree an imponderable. That aside, to my mind, the largest tractable source of uncertainty that remains is about how best to extrapolate from these results from historical, untreated settings to modern, treated settings. It is not obvious that a child measured at WHZ = –2 in a historical setting is a good proxy for a child whose WHZ today is lifted by CMAM from –3 to –2. One virtue of the MIG model is that its microtheory could allow a more sophisticated simulation of treatment, e.g., as a sustained increase in health trajectory. Such simulation would require

[2] Countries are Burkina Faso, Chad, the DRC, Ethiopia, Kenya, Mali, Niger, and Nigeria. Average incorporates survey weights.

data on the distribution of anthropometric trajectories among children deemed eligible for treatment today.

Table 1. One-year mortality ratio estimates after random-effects meta-analysis

Model	WHZ = –3.5 vs. –0.5	MUAC = 11 vs. 14	WHZ = –2.75 vs. –1.75	MUAC = 11.75 vs. 12.75
MIG	5.61 [2.58, 12.18]	4.56 [1.96, 10.61]	1.81 [1.38, 2.36]	1.73 [1.32, 2.255]
Probit	3.75 [2.22, 6.31]	3.51 [1.84, 6.70]	1.58 [1.31, 1.92]	1.56 [1.22, 1.99]

Note: 95% confidence intervals in brackets.

1 Previous literature

Olofin et al. (2013) continues a lineage of research that pools data previously collected in multiple settings.

1.1 Pelletier et al. (1994)

Pelletier et al. (1994) gathers summary statistics on eight data sets like those in Olofin et al.: three from Bangladesh, and one each from India, Indonesia, Papua New Guinea, Tanzania, and Malawi. From the published reports, Pelletier et al. extracts sample sizes and death counts. The study disaggregates the statistics according to a once-common anthropometric classification, with groups demarcated at 60%, 70%, and 80% of the standard median weight-for-age. Those boundaries are about one z score unit lower than those in more recent studies (i.e., $z \approx -4, -3, -2$).[3]

Pooling across the eight studies, Pelletier et al., Table 3, finds a death rate 8.4 times higher among the most underweight children (weight-for-age $z \lesssim -4$ on the modern WHO standard) as compared to the least underweight ($z \gtrsim -1.75$). For lack of access to the underlying microdata, Pelletier et al. cannot perform survival analysis or other modeling of it. The regressions in the study are in a sense meta-regressions, as their unit of observation is a weight-for-age group within a given data set.

1.2 Fishman et al. (2004)

Fishman et al. (2004) is strongly influenced by Pelletier et al. The underlying studies now number ten and come from ten countries: Bangladesh, Ghana, Guinea-Bissau, India, Indonesia, Nepal, Pakistan, Philippines, Senegal, Sudan. Fishman et al. extends the earlier work by collecting the microdata. However, Fishman et al. does not directly model the data either. Instead, it computes the aggregates needed in order to follow the methods of Pelletier et al., while applying a more modern anthropometric classification, namely, weight-for-age with z score boundaries at -3, -2, and -1 under the WHO/NCHS 1977 standard.

Fishman et al., Table 2.6, reports a risk ratio of 8.72 between most- and least-underweight groups for all-cause mortality. Disaggregating by cause of death produces ratios ranging from 5.22 for measles to 12.50 for diarrhea.

1.3 Black et al. (2008)

Black et al. (2008) continues the development from Fishman et al.—though precisely how is hard to determine because the description of the relevant methods is a few sentences (p. 247). For reasons not stated, Black et al. does not use data from Indonesia or Sudan.[4] The paper says that "generalised linear mixed models were used," but it does not say which ones; a good guess, since odds ratios are reported, is a

[3] Six of the eight studies reference the WHO/NCHS 1977 medians (Pelletier et al. 1994, Table 1, note 1). The resulting 60%, 70%, and 80% threshold correspond to $z = -4.05, -2.9,$ and -1.8 on the WHO 2006 standard now in use, averaging over age and gender values. The remaining two studies refer to the "Harvard" standard, for which the resulting modern z values are $-4.0, -2.8,$ and -1.7. The WHO/NCHS standard is at web.archive.org/20050930101846/http://www.who.int/nutgrowthdb/reference/Weight_for_age.csv. The Harvard standard is at who.int/iris/bitstream/handle/10665/41780/WHO_MONO_53_(part4).pdf#page=46.

[4] In section 2 below, I explain why, despite obtaining the Indonesia data, I do not use it.

logit model with study-level random effects. The study may also be the first to exploit access to the microdata by modeling it directly. The terse mention of GLMs seems to imply otherwise—to align the paper with previous studies, which fit GLMs only to aggregates.[5] Yet Olofin et al., which shares two authors with Black et al., states that the latter fits GLMs "using individual-level data."

Other novelties in Black et al. include the use of the WHO 2006 growth standards to classify subjects, and the extension of the analysis from weight-for-age to height-for-age and weight-for-height.

The results for weight-for-age are similar to those in Fishman et al., though generally a bit smaller (Fishman et al., Table 2.6; Black et al., Table 2, first panel). The odds ratio associated with weight-for-age $z < -3$ versus $z > -1$ is 9.7.

1.4 Olofin et al. (2013)

Olofin et al. (2013) departs from previous work in several ways. It returns to the full 10 data sets in Fishman et al. It applies a method from survival analysis proper, meaning that it incorporates information not only about whether children die, but when. In particular, it runs Cox proportional hazards regressions. Where data availability allows, Olofin et al. incorporates aggressive control sets. And in order to blend results from different control sets and from different settings, it performs DerSimonian and Laird (1986) random effects meta-analysis. Although the results are now hazard ratios instead of cumulative mortality ratios, the pattern of results remains familiar. Under the WHO 2006 standard, the hazard ratio for children with weight-for-age z below –3 is 9.40. The corresponding hazard ratio for WHZ < -3 versus WHZ > -1 is 11.63.

2 Data

GiveWell requested data from the corresponding authors of Olofin et al. (2013) as well as of the underlying studies. We obtained data for five of the ten: Adair et al. (1993) from the Philippines, Garenne et al. (2000) from Senegal, Katz et al. (1989) from Indonesia, Mølbak et al. (1992) from Guinea-Bissau, and West et al. (1991) from Nepal.[6] We also obtained the data for one study not in Olofin et al., that of Van den Broeck, Eeckels, and Vuylsteke (1993), which took place in what is now the Democratic Republic of Congo. That data set is in the public archive for Schwinger et al. (2019).[7]

All six data sets supply repeated anthropometric measurement in a community-based cohort of young children, along with information on the occurrence and timing of death. The data sets differ in most other respects—sample size; which measurements were taken (height/length, weight, MUAC); precision of dates; richness of information on morbidity, cause of death, access to clean water, and demographics; and whether a randomized trial of vitamin A supplementation was ongoing within the sample.

One limitation became apparent in the Indonesia data: dates of death are not recorded, only dates of the first researcher visit after death.[8] Survival modeling ordinarily demands precise information on timing of failures. The corresponding author for Olofin et al., Goodarz Danaei, was uncertain about how this issue was handled in that analysis, and suggested generating random dates between the visits bracketing death. Rather than engaging in ad hoc imputation, I applied interval-censored variants of the Cox and MIG

[5] Pelletier et al. (1994) "estimated with an unweighted regression (random effects model)." Fishman et al. "followed the procedures of Pelletier et al. (1993, 1994, 1995) and used the SAS Proc Mixed program" (p. 56).
[6] The Adair et al. (1993) data are online at dataverse.unc.edu/dataverse/cebu.
[7] search.nsd.no/en/study/NSD2709.
[8] Confirmed in email from Joanne Katz, Johns Hopkins University, November 17, 2021.

models.[9] However, these often did not converge. Perhaps this is to be expected since the failure dates are completely censored within a spell. For this reason, Indonesia data is not used with the Cox and MIG models.

All WHZ scores are recomputed from primary data according to the WHO 2006 standard (WHO 2009, pp. 302–03), using posted reference tables.[10] The WHO and UNICEF (2009) define non-edematous SAM for children 6–59 months old as having WHZ < –3 or MUAC < 11.5 cm. The standard thresholds for MAM are one unit higher: WHZ = –2 and MUAC = 12.5cm. ("Global Acute Malnutrition," or GAM, includes SAM and MAM.)

Through the lens of survival analysis, the raw data define multiple episodes (spells) per child. Each episode begins with a measurement of WHZ or MUAC, the two SAM-diagnostic metrics. Each ends with the next measurement, usually 2–6 months later; or with death before the next researcher visit; or with a survival status report from an endline mortality survey. Together, the six data sets contain 89,142 episodes beginning with WHZ measurement between ages of 6 and 60 months, and 135,799 beginning with MUAC measurement; the two groups overlap since sometimes both metrics were collected. 22,372 children figured in the WHZ episodes and 40,253 in the MUAC episodes. 978 of the children were measured at least once with what are now SAM-prognostic values of WHZ (< –3) and 3298 with MUAC-SAM (MUAC < 11.5cm). The 1,411 episodes following those WHZ measurements totaled 855 child-years in length and led to 99 deaths; for MUAC 5,114 episodes totaling 2,360 child-years led to 235 deaths.

See Table 2 for details. Among the patterns in the table:

- Deaths/episode and deaths/year indeed decline as anthropometric indicators increase.
- Death rates in corresponding WHZ and MUAC categories are often reasonably similar[11], the largest exception being the SAM category in the DRC, where the SAM mortality was approximately 10 times higher per episode or unit of time when indicated by WHZ than by MUAC.
- The MUAC statistics are dominated by the large Nepal sample.
- Average episode length is longest in Senegal. There, the researchers conducted a mortality follow-up survey at the end of 1989, five years after the initial two years of semimonthly measurement.
- Children often changed categories from measurement to measurement. As a result, the sums of the numbers of children appearing in each category at some point exceed the values in the Total columns.

Figure 1 renders the entire data set, in a sense. Children's histories in WHZ or MUAC are plotted, and deaths are marked at the point of last measurement. We see that anthropometry can change quickly, perhaps because of measurement error, but tends to improve with age. In some settings, death rates clearly fall with age.

[9] For Cox regression, the method was that of Turnbull (1976), as implemented in the Stata command stintcox. For the inverse Gaussian model, the likelihood of death over the full spell is computed as the complement of the survival probability, in my Stata implementation of the model.

[10] github.com/WorldHealthOrganization/anthro/tree/master/data–raw/growthstandards. Only the Nepal and Senegal data indicate whether height or length was measured. Elsewhere, weight–for–length references are used for age < 2 years and weight–for–height references otherwise.

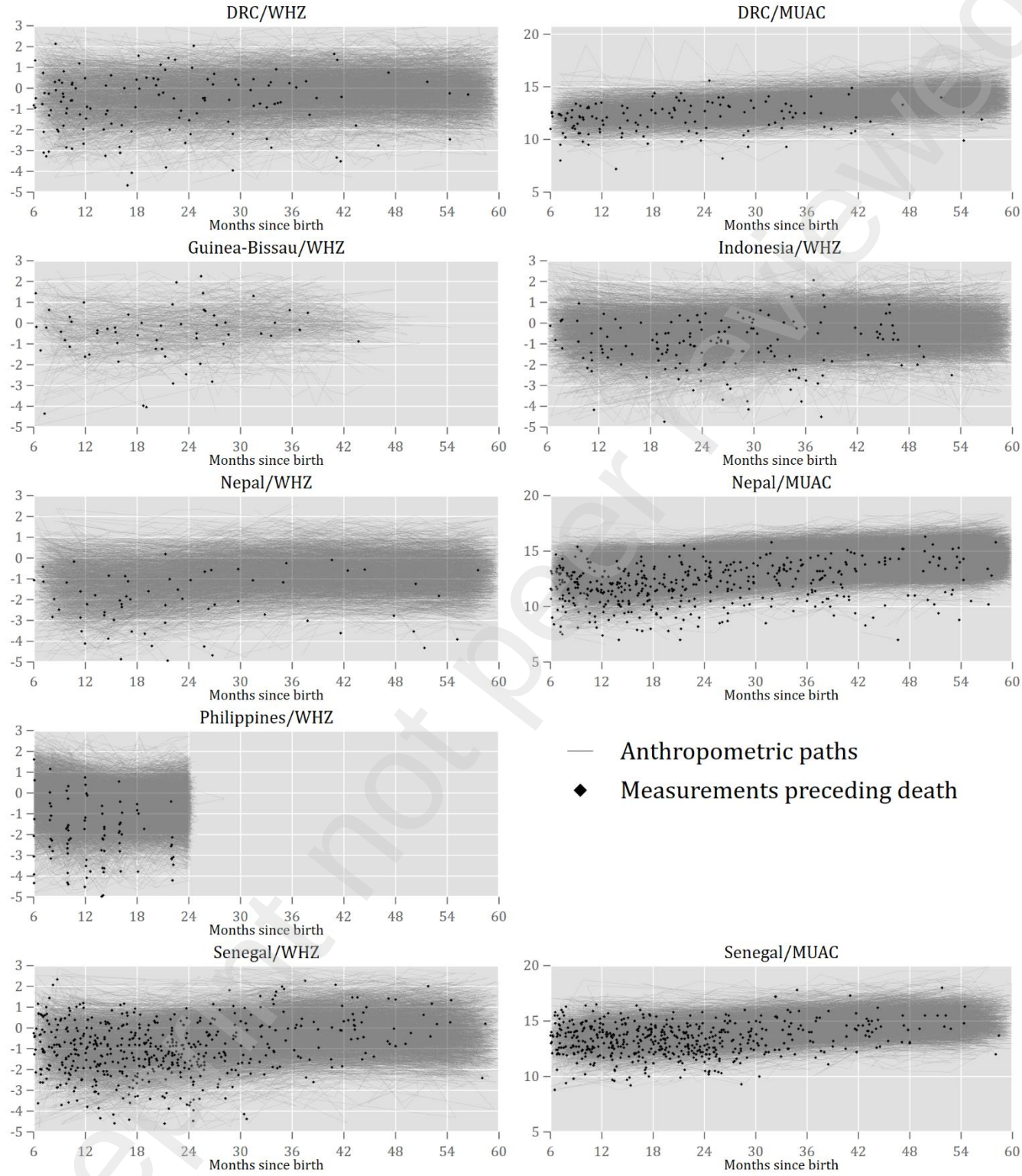
[11] This does not mean that WHZ and MUAC criteria are identifying the same children (Grellety and Golden 2018).

Table 2. Numbers of subjects, episodes, and deaths by malnourishment metric

	Weight-for-height *z* score (WHZ)					Mid-upper-arm circumference (MUAC, cm)				
	< −3	−3 to −2	−2 to −1	> −1	Total	< 11.5	11.5–12.5	12.5–13.5	> 13.5	Total
Democratic Republic of Congo (Van den Broeck et al. 1993)										
Subjects	113	470	1519	4146	4587	829	1858	2562	2542	4598
Episodes	137	694	2984	14160	17975	1490	3779	5724	7302	18295
Measurement age (months)	23.0	25.4	29.6	33.7	32.6	20.6	24.5	31.7	40.0	32.6
Years of risk	58.2	333	1526	7411	9328	705.8	1832	2951	3987	9475
Months/episode	5.1	5.8	6.1	6.3	6.2	5.7	5.8	6.2	6.6	6.2
Deaths	10	15	21	79	125	43	29	37	19	128
Deaths/episode	0.073	0.022	0.007	0.006	0.007	0.029	0.008	0.006	0.003	0.007
Deaths/year of risk	0.172	0.045	0.014	0.011	0.013	0.061	0.016	0.013	0.005	0.014
Guinea-Bissau (Mølbak et al. 1992)										
Subjects	23	80	230	612	664					
Episodes	29	131	467	2647	3274					
Measurement age (months)	21.6	20.8	21.5	24.7	24.1					
Years of risk	8.9	49	186	1033	1277					
Months/episode	3.7	4.5	4.8	4.7	4.7					
Deaths	3	3	9	42	57					
Deaths/episode	0.103	0.023	0.019	0.016	0.017					
Deaths/year of risk	0.336	0.061	0.048	0.041	0.045					
Indonesia (Katz et al. 1989)										
Subjects	115	419	1550	3514	3900					
Episodes	139	619	2930	11398	15086					
Measurement age (months)	25.7	25.5	30.0	33.9	32.7					
Years of risk	40	170	817	3245	4272					
Months/episode	3.5	3.3	3.3	3.4	3.4					
Deaths	10	20	43	80	153					
Deaths/episode	0.072	0.032	0.015	0.007	0.010					
Deaths/year of risk	0.247	0.118	0.053	0.025	0.036					
Nepal (West et al. 1991)										
Subjects	307	1095	2695	3772	5309	2241	6355	14039	22655	30511
Episodes	449	1841	5481	10538	18309	3348	10121	26478	64919	104866
Measurement age (months)	20.2	26.5	32.0	35.3	33.1	16.5	20.4	27.4	37.9	32.9
Years of risk	146.1	610	1811	3461	6029	1094.2	3349	8787	21304	34535
Months/episode	3.9	4.0	4.0	3.9	4.0	3.9	4.0	4.0	3.9	4.0
Deaths	16	16	18	13	63	146	89	87	89	411
Deaths/episode	0.036	0.009	0.003	0.001	0.003	0.044	0.009	0.003	0.001	0.004
Deaths/year of risk	0.109	0.026	0.010	0.004	0.010	0.133	0.027	0.010	0.004	0.012
Philippines (Adair et al. 1993)										
Subjects	202	757	1760	2444	2771					
Episodes	408	1707	5896	13877	21888					
Measurement age (months)	13.8	14.3	14.5	14.1	14.2					
Years of risk	66.7	289	1002	2356	3714					
Months/episode	2.0	2.0	2.0	2.0	2.0					
Deaths	24	20	18	21	83					
Deaths/episode	0.059	0.012	0.003	0.002	0.004					
Deaths/year of risk	0.360	0.069	0.018	0.009	0.022					
Senegal (Garenne et al. 2000)										
Subjects	218	673	1799	4094	5141	228	673	1503	4252	5144
Episodes	249	825	2635	8901	12610	276	820	2082	9460	12638
Measurement age (months)	19.5	21.3	27.4	34.3	31.7	18.6	18.7	23.5	35.1	31.7
Years of risk	535.0	1832	5981	20551	28899	559.6	1758	4518	22138	28973
Months/episode	25.8	26.7	27.2	27.7	27.5	24.3	25.7	26.0	28.1	27.5
Deaths	36	71	151	275	533	46	95	121	275	537
Deaths/episode	0.145	0.086	0.057	0.031	0.042	0.167	0.116	0.058	0.029	0.042
Deaths/year of risk	0.067	0.039	0.025	0.013	0.018	0.082	0.054	0.027	0.012	0.019
Total										
Subjects	978	3494	9553	18582	22372	3298	8886	18104	29449	40253
Episodes	1411	5817	20393	61521	89142	5114	14720	34284	81681	135799
Measurement age (months)	19.1	21.8	25.5	29.3	27.8	17.8	21.4	27.9	37.7	32.7
Years of risk	855.4	3283	11323	38058	53520	2359.6	6939	16256	47429	72983
Months/episode	7.3	6.8	6.7	7.4	7.2	5.5	5.7	5.7	7.0	6.4
Deaths	99	145	260	510	1014	235	213	245	383	1076
Deaths/episode	0.070	0.025	0.013	0.008	0.011	0.046	0.014	0.007	0.005	0.008
Deaths/year of risk	0.116	0.044	0.023	0.013	0.019	0.100	0.031	0.015	0.008	0.015

Notes: Samples restricted to episodes starting at age 6–60 months. Episodes begin with measurement and end with subsequent measurement, death, or survival report from endline mortality survey. Children who changed anthropometric groups over time are counted in more than one column. Deaths/year-of-risk figures slightly upward-biased because death curtails risk exposure periods.

Figure 1. Weight-for-height z (WHZ) and mid-upper-arm circumference (MUAC) paths, with deaths, all data sets



3 Analysis in narrative form

This section works to develop an analysis of the relationship between anthropometry and mortality in incremental fashion: starting with some basic facts about the data; explaining concerns that complicate

interpretation and extrapolation to modern treatment contexts; and sequentially introducing complications motivated by those concerns.

3.1 Plotting deaths/episode

To start, Figure 2 organizes information on the most basic outcome measure in the data set, the fraction of episodes ending in death. Each row in the figure shows statistics for one data set and anthropometric categorization, such as Nepal and WHZ. As is standard, the WHZ spectrum is broken into four ranges with cut points at −3, −2, and −1. Correspondingly, the MUAC range is cut at 11.5, 12.5, and 13.5. For both metrics, the first cut point defines SAM and the second MAM.

The left pane plots the rates from Table 2 along with 90% confidence ranges. Successive numbers are placed alternately above or below the data points to reduce collisions. WHZ and MUAC rows for same country are grouped, where applicable. Countries are sorted vertically by the mortality rate associated with SAM-WHZ —from 0.036 deaths/episode in Nepal to 0.14/episode in Senegal.

Some of the ratios between the numbers in the left pane appear on the right: in each row, the risk rates for the three less-healthy categories are divided by that for the healthiest one.

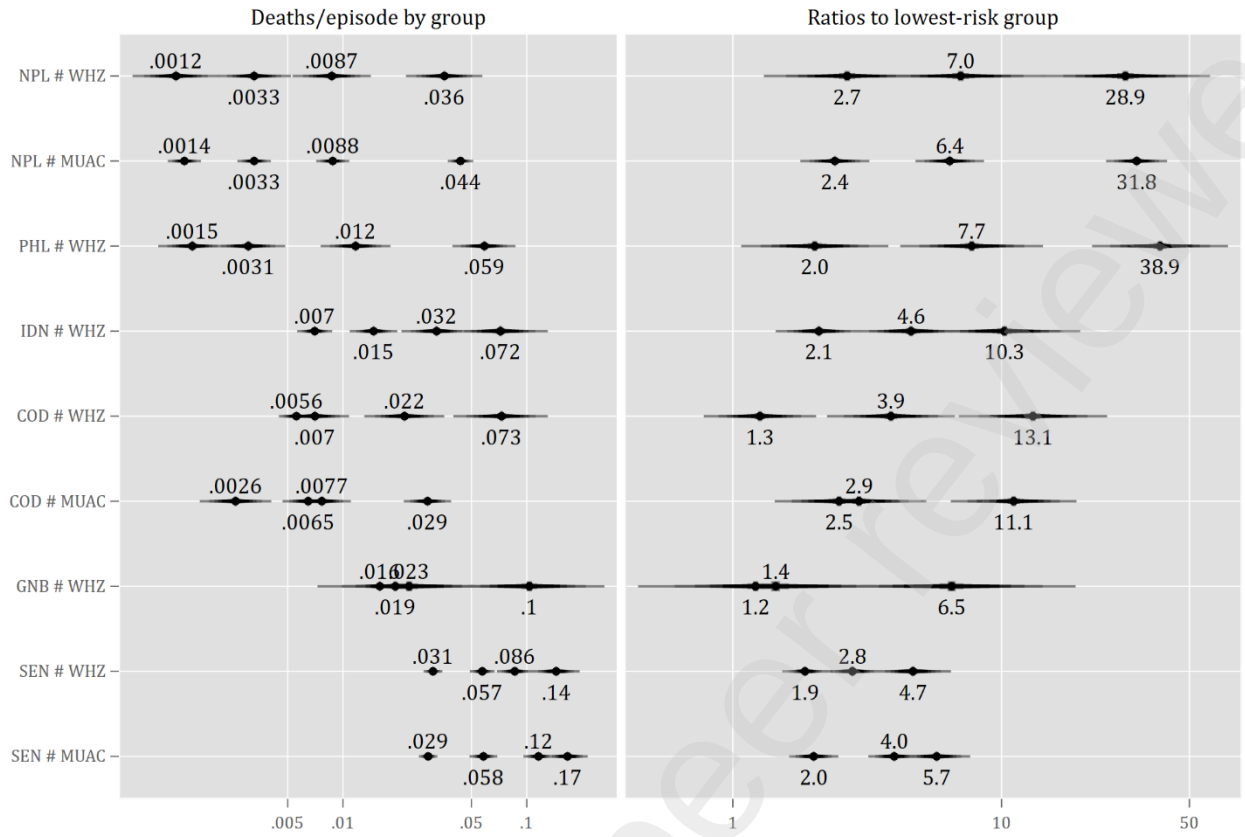
A few patterns emerge in the figure:

- Mortality rates per episode tend to rise and fall together across anthropometric categories. For example, the Senegal data are the locus for the highest per-episode mortality in every WHZ category.
- Where mortality is higher, mortality ratios are *lower*. On the logarithmic scales of the figure, the dots are closer together at the bottom in the left pane and closer to the left in the right pane.
- Where mortality is higher, overall mortality *differences* are *higher*. Precisely because of the logarithmic scales, this fact is less obvious. But examining the numbers will confirm. The spread across the range for Nepal-WHZ, 0.036 – 0.0012 = 0.035, is much smaller than that in the high-mortality setting of Senegal, 0.17 – 0.029 = 0.14.

The second and third facts plausibly suggest that in general the factors elevating child mortality in some settings partly interact with malnourishment to cause death, and partly do not. To the extent they do, they would increase risk differences and preserve risk ratios—e.g., if malaria causes death in strict proportion to malnutrition. To the extent that factors such as vehicular accident risk do *not* interact with malnourishment, they would raise risk in all anthropometric categories and reduce ratios across them.

But part of the story may be about differences in how frequently researchers measured children for each of the data sets, as will be explored below. In the meantime, I will follow convention and focus on risk ratios, as in the right pane of Figure 2.

Figure 2. Deaths/episode by anthropometric group, and ratios thereof to rates for healthiest group



Notes: Each row shows statistics for one data set and anthropometric categorization. The left pane shows the fraction of episodes leading to death, broken out by the same categorizations as in Table 2. The right pane shows the ratios of those risk rates to the risk rate in the healthiest group (WHZ > −1 or MUAC > 13.5). Degrees of confidence are indicated using standard errors clustered by subject. NPL = Nepal; PHL = Philippines; IDN = Indonesia; COD = Democratic Republic of Congo; GNB = Guinea-Bissau; SEN = Senegal.

3.2 Complications

Even if there were a compelling story favoring differences or ratios, the question of how best to extrapolate from untreated, historical settings to prospective treatment settings would be complicated. In a sense the question is one of interpretation: what do these numbers tell us about what we care about? Complications in interpretation tend to lead to complications in method, as they motivate mathematical elaborations to reduce, say, reverse causation. Here are complications that GiveWell has voiced or that I have arrived at in the course of analysis:

1. *Confounding.* Low nutritional status may be a correlate of non-nutritional causes of death, as distinct from a proximate cause of death. This could lead to overestimation of the causal impact of malnourishment on mortality.
2. *Nonrepresentativeness.* Children in these data sets could be older on average than the ones offered CMAM today, and older children with SAM are less likely to die (a fact confirmed below).
3. *Dummies are dumb?*[12] Representing nutritional status with dummies—0/1 indicators for broad ranges such as WHZ < −3—throws away information and results in a coarse representation of the statistical patterns. Interpreted mechanically, risk rates such as those in Figure 2 imply that moving a hypothetical child from WHZ = −4 to WHZ = −2.1 does exactly as much as good as moving from one WHZ = −3.01 to WHZ = −2.99, since both changes would formally shift the child from SAM to MAM. That

[12] Turn of phrase aside, the problem can also be cast as *too few* dummies. Finer categorization might suffice.

example is contrived, but for purposes of impact estimation GiveWell's Stephan Guyenet suggested I assume SAM children see their WHZ increase by about 1.5 points (Table 8), and whether those increases in hypothetical children happen to cross one or two category boundaries could significantly affect the estimated risk reduction.

4. *Not incorporating timing information*. Suppose a study reveals that after getting a positive result on some test for lung cancer, 10% of patients die within 6 months. How would we estimate the 12-month risk? A natural guess would be that doubling the period doubles the risk. Indeed, if it were further revealed that the deaths in the study occurred evenly across the 6 months, that would bolster the guess. On the other hand, if it turned out that *all* of the deaths occurred within a week of the test, that would undermine the guess. Evidently patients who survived that initial, fraught period were safer thereafter.

   In addition, timing information can give insight into causality—though I am uncertain about how to apply this idea in the present context. If statisticians find a burst of deaths in the days following car accidents, that tells a compelling causal story. If they merely find that people who are in car accidents die more in the following decade, it becomes more plausible that propensity for collisions mostly indicates other risk factors.

   The data sets at hand do tell us not only whether deaths occurred within set periods, but when. The simple averages in Table 2 and Figure 2 discard this information, and may lead to worse extrapolations to other time frames. Survival analysis methods would allow us to incorporate timing, a standard example being the Cox proportional hazards model used in Olofin et al.

   Caveats: As mentioned, the Indonesia data set lacks timing information. And timing information in other data sets could be inaccurate, especially if it is based on parents' recall of events that happened months or years back.[13]

5. *Risk ratios vary over time.* If, as in the lung cancer hypothetical, the risk of death after SAM diagnosis varies over time, then the death rates on the left side of Figure 2 are not directly comparable across countries. The low risks in the Philippines, for example, come from tracking kids for just two months before the next measurement while the high death rates in Senegal are for episodes lasting more than two years on average.

   One might hold out hope that the risk *ratios* in the right pane of the figure are comparable despite differences in follow-up frequency. Indeed, the proportional hazards model is called that precisely because it assumes as much.[14] For example, perhaps after SAM diagnosis the moment-to-moment risk of death (hazard) first rises for a while and then falls (among survivors). And perhaps it does the same in children who are measured as above SAM thresholds, if with lower amplitude (though that sounds unlikely). Then while the risks varied, the risk ratios would not. The prospect is attractive because it can simplify analysis. But the assumption of proportional hazards is normally made for mathematical

[13] "Age is the most problematic of the variables. In most Western cultures it is rounded down to the nearest completed month, in other cultures age is rounded up to the next highest month. In most developing countries there is no birth registration and birthdays are not celebrated; consequently, as a child ages the actual age becomes increasingly vague in the memory of the mother. Age is then approximated by using a calendar of local memorable events and the mother asked to remember to which event the birth most closely approximated. These calendars are normally quite crude and are not capable of identifying the actual month of birth. The questioning takes some time and is often administered in a perfunctory way, particularly if the survey team is tired or asked to complete excessive interviews in a single day." (Grellety and Golden 2016, p. 5) Though this is about time of birth, much of it applies to time of death.

[14] Technically, the proportional hazards assumption is that risk ratios are the same at each *age*, not each time since last measurement. However, in the special case where children are measured at the same ages, the distinction evaporates, and the proportional hazards assumption implies risk ratios are constant with respect to time since measurement.

convenience, not biological realism.[15] In our case, it seems unlikely that the time profile of death among well-nourished children would be a scaled copy of that for SAM children.

Another way of describing the situation: the statistics in Figure 2 are about *different risks*. Since the Philippines data collectors measured children every two months, at any given moment in those 3,714 child-years of risk exposure (see Table 2) a child was on average measured about one month previously. By the same reasoning, the Nepal statistics shed light on the risk of death conditional on having been diagnosed with SAM a two months ago. And so on. There is no strong *a priori* reason to believe that cross-group ratios of different risks are the same.

In responding to these concerns in the analysis, I will start with those that require the least methodological complication, in order to maintain accessibility.

3.3 Risk ratios vary over time

I see two main approaches to increasing the comparability of results from data sets with various follow-up frequencies. One, complicated, is to devise a mathematical model that explicitly represents how risks evolve over time, then calibrate it to the data. The other is to homogenize the data so that all subjects are followed the same amount of time. This throws away most of the information in the data set, but is more transparent.

Sticking with the simple approach here, I reorganize the data so as to represent each child with a single spell starting with the first measurement after reaching six months of age and running one year. Children are only retained in the data if they are followed for at least that long, or die before then. Subsequent measurements within these 12 months are ignored, as is any information pertaining to events happening after. I start at six months because it is the earliest age at which children can be diagnosed with GAM under WHO guidelines, and because malnutrition most endangers the youngest children. I follow children for a year because in discussions with GiveWell staff, cumulative one-year mortality has been favored as a benchmark, and because 12-month timing meshes well with the approximate 2-, 3-, 4-, and 6-month cadences in the raw data.[16]

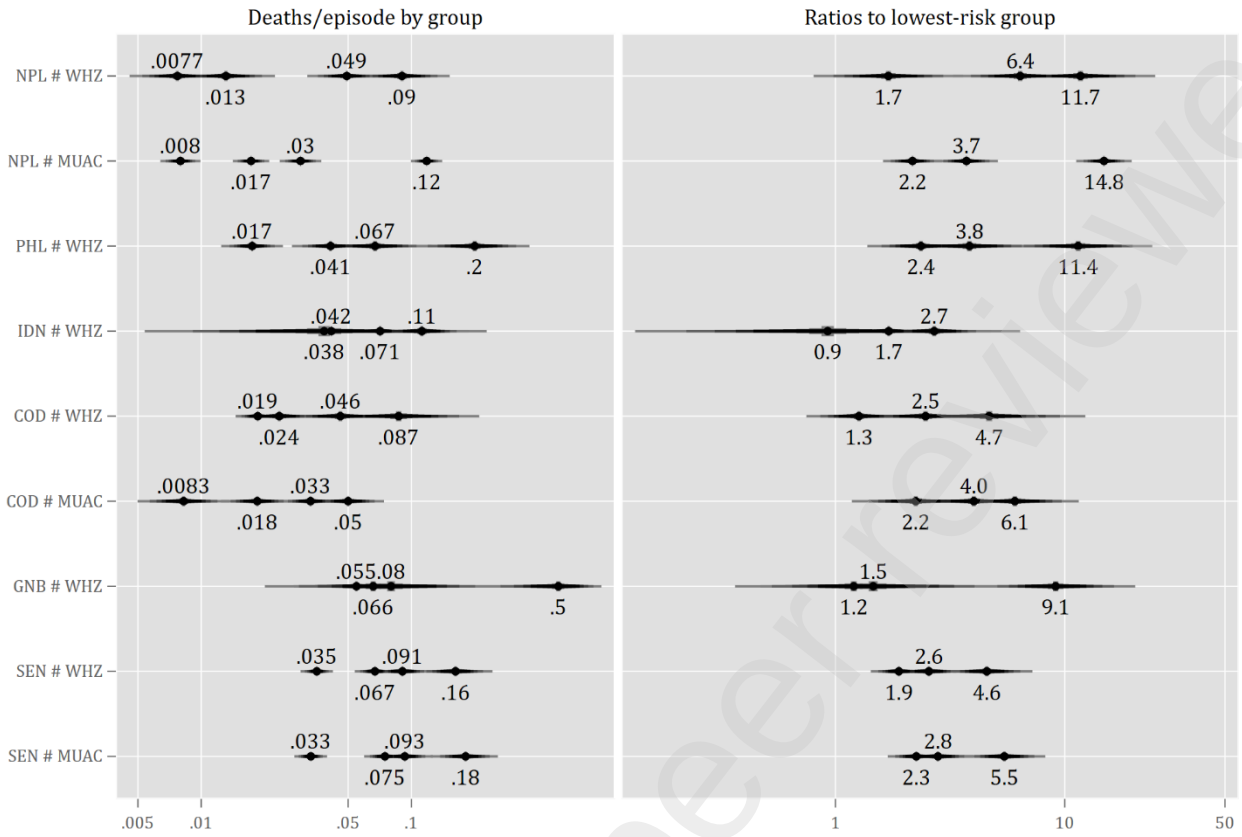
Figure 3 shows the mortality averages and ratios in these "single-spell" data sets. I note that:

- Cumulative risks (left pane) all increase from the previous figure, which essentially they must since the subjects are exposed to risk longer, except in Senegal.
- Just as in the cross-country comparisons within Figure 2, the increased risks from the data restructuring tends go hand in hand with lower risk *ratios*. The highest ratios in Figure 2, for Nepal and the Philippines, are replaced with ones half as much or less. This suggests that the proportional hazards assumption does not hold in all the data sets.
- After the single-spell restructuring, risk ratios are less varied. It appears that taking steps to address timing effects can indeed produce results that are more consistent across data sets.

[15] "In general, there are (at least) two forces that might lead to diminishing the effect of a covariate, namely ageing of measurements and frailty….Hence one should clearly be cautious in assuming a constant effect of a covariate over time as is done in the proportional hazards setting. In regression modeling for survival data, there is typically too little emphasis on changes in effects over time, and hence one gets an unnecessarily limited understanding of the process. There is in general no reason why effects should be constant over time." (Aalen, Borgan, and Gjessing 2008, p. 163)

[16] The lack of timing information in the Indonesia data set makes it impossible to estimate whether deaths in spells bracketing the one-year mark occurred before or after that moment, thus whether the child survived a full year.

Figure 3. Deaths/episode by anthropometric group, and ratios thereof to rates for healthiest group, one-year spells



3.4 Nonrepresentativeness (in age)

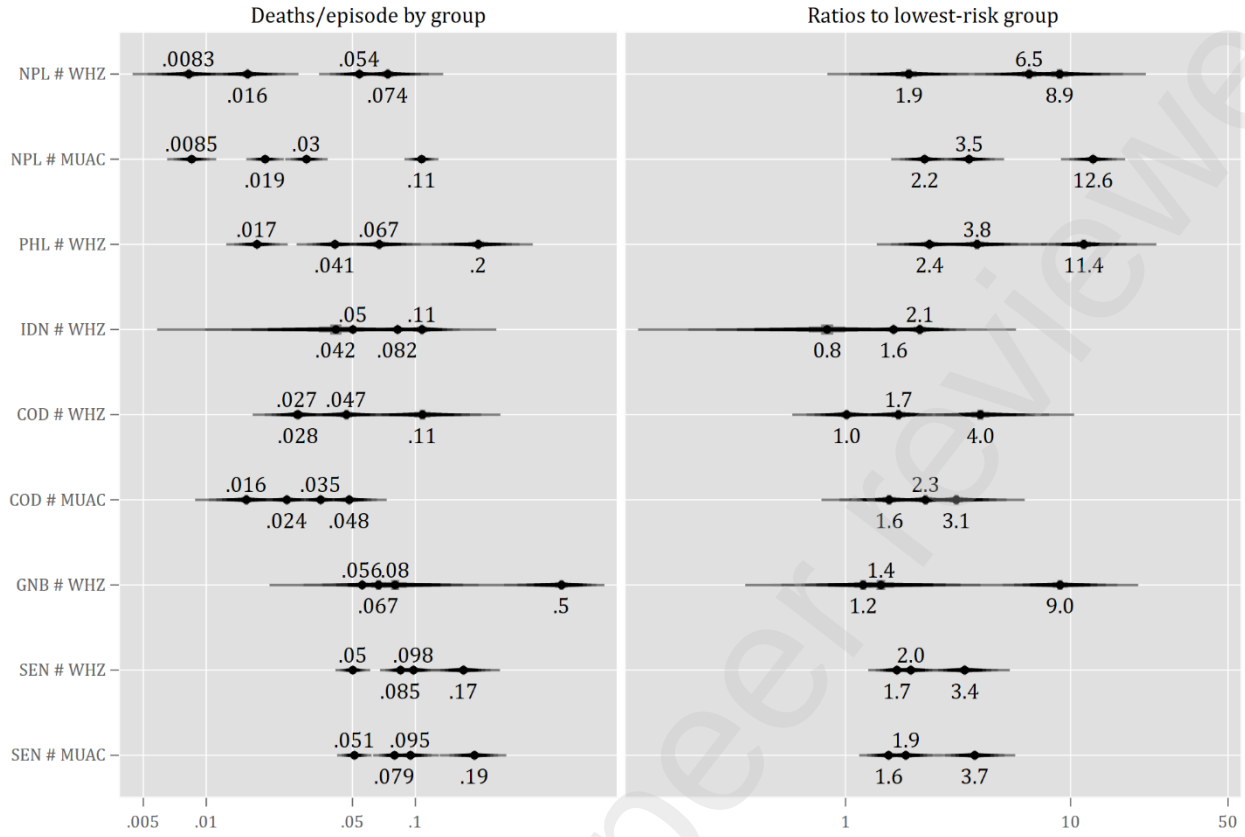
Those one-year spells begin at an average age of 23.6 months for WHZ and 25.8 months for MUAC. But the children recruited into CMAM today might be younger. In a recent field trial in Burkina Faso for children 6–59 months, the average age of those recruited for treatment (having oedema or MUAC < 12.5cm) was just 14.9 months (Daures et al. 2020, Table 2).

To increase comparability on age, I narrow the samples behind the previous figure. Episodes are retained only if they begin before age 36 months, instead of 60 months. This reduces average age at measurement to 15.7 months for WHZ and 18.4 for MUAC.

The impacts of the change on risks are subtle. See Figure 4. The major difference I see between Figure 3 and Figure 4 is that death rates among the well-nourished rise a bit, causing the risk ratios, which take them as a reference, to fall slightly.

Because of the modest effect and the further loss of sample, I will dispense with this restriction in what follows.

Figure 4. Deaths/episode by anthropometric group, and ratios thereof to rates for healthiest group, one-year spells starting before age 36 months



3.5 Confounding

The remaining concerns introduce enough complexity that they can only be addressed by imposing further assumptions, which brings us to statistical modeling. For example, to address confounding we may wish to "control for" observed socioeconomic factors. This usually requires assuming that the factors have a relationship with mortality risk whose mathematical structure is known, and thus that we are only ignorant of its parameters. The parameters are then estimated by fitting the model to the data.

To maintain continuity with the analysis to this point, I keep the single-episode-per-child data set but switch to a probit model, which is standard for binary outcomes. A probit model for death in each of these samples, whose regressor set consists solely of dummies for the four WHZ or MUAC ranges, leads to probability predictions identical to those in Figure 3—so those are not plotted again.

But on that foundation, we can build a more elaborate structure. To the probit regressions I add controls for age, gender, and, where defined, the Olofin et al. control sets, which are described here in Table 3. (For lack of data, Olofin construct no controls for the DRC and Senegal data sets.) After fitting the model to each data set, I generate predictions of mortality risk for each WHZ or MUAC category by setting all other variables to their sample means. I drop the data for Indonesia because the software was unable to generate probability predictions, evidently under the combined stresses of sample shrinkage (to one-year spells) and addition of many controls. These specification choices are meant to minimize my own exercise of discretion, by accepting the most obvious defaults.

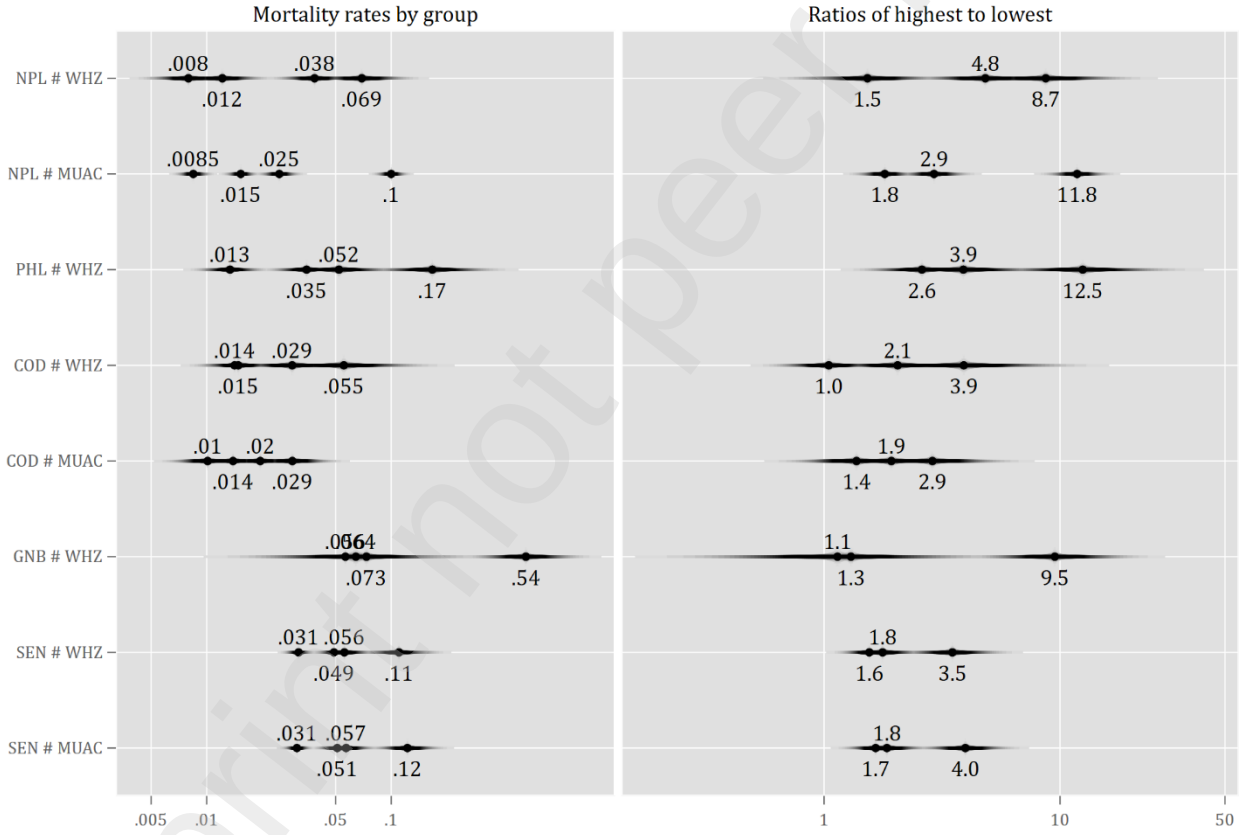
The results appear in the familiar format in Figure 5. The departures from Figure 3 are modest, and I see no strong pattern in them.

Table 3. Olofin et al. description of control sets copied in present analysis

Country	Covariates
Guinea-Bissau	Ethnic group
Indonesia	Mother's education, mother's age, household drinking water source, family size
Nepal	Household assets (household ownership of: bari, khet; number of household owned: bicycles, cattle, goats, radios; materials used for house walls and roof), mother's education, mother's age, household latrine ownership, household caste
Philippines	Sanitation variables (i.e., animals kept under the house, domestic animals kept inside the house, condition of food cooking area, condition of food storage area), whether or not the child's mother works for pay, source of baby's drinking water, household water supply (availability of: boreholes, dug wells, piped supply, springs)

Source: Olofin et al. (2013), Table S1.

Figure 5. Predicted deaths/episode by anthropometric group, and ratios thereof to rates for healthiest group, one-year spells, probit model with Olofin et al. controls where available



3.6 Smarter than dummies?

The standard alternative to a set of dummies is a set of polynomial terms in the underlying continuous variable—typically first- or second-order, meaning linear or quadratic. Polynomial terms have the advantage that to the extent that they capture the functional relationship between dependent and independent variables, they can allow more precise characterization of fit. They could more properly distinguish between a move from WHZ = –4 to WHZ = –2.1 and one from –3.01 to –2.99.

Figure 6 documents an exploration of this option in the context of the single-spell data sets. The figure contains one plot for each row in the previous figures.[17] Each plot in turn has:

- A horizontal axis for WHZ (from –4 to 0) or MUAC (9.5 to 14.5cm).
- A vertical axis for average mortality.
- Black dots for each child, showing the child's initial anthropometry and subsequent survival or death. The deaths appear along the top and the survivals along the bottom. Noise is added to the placement of the dots to better convey the density of the data.
- Grey, wiggly lines that are smoothed fits (moving averages) of the 1's and 0's; they represent death and survival rates as a function of WHZ or MUAC.
- White bands for the 95% ranges of the smoothed fits.
- Blue step functions that show the mortality rates predicted in each anthropometric group by a fitted probit model using the standard four dummies and no controls. The heights of these steps are the numbers in the left pane of Figure 3. They are the average numbers of deaths among the children whose dots are directly above or below them.
- Teal curves showing the mortality predictions from a fitted probit model that replaces the four dummies with a single, continuous WHZ or MUAC variable.
- Gold curves that do the same for a quadratic probit model.

One can see that smoothed fits often behave strangely near the edges, especially if the data is sparse there. The plot for Indonesia, for example, shows *falling* risk as WHZ declines below −3, because there happen to be almost no deaths in that range in the single-spell data. Not much should be made of such end effects.

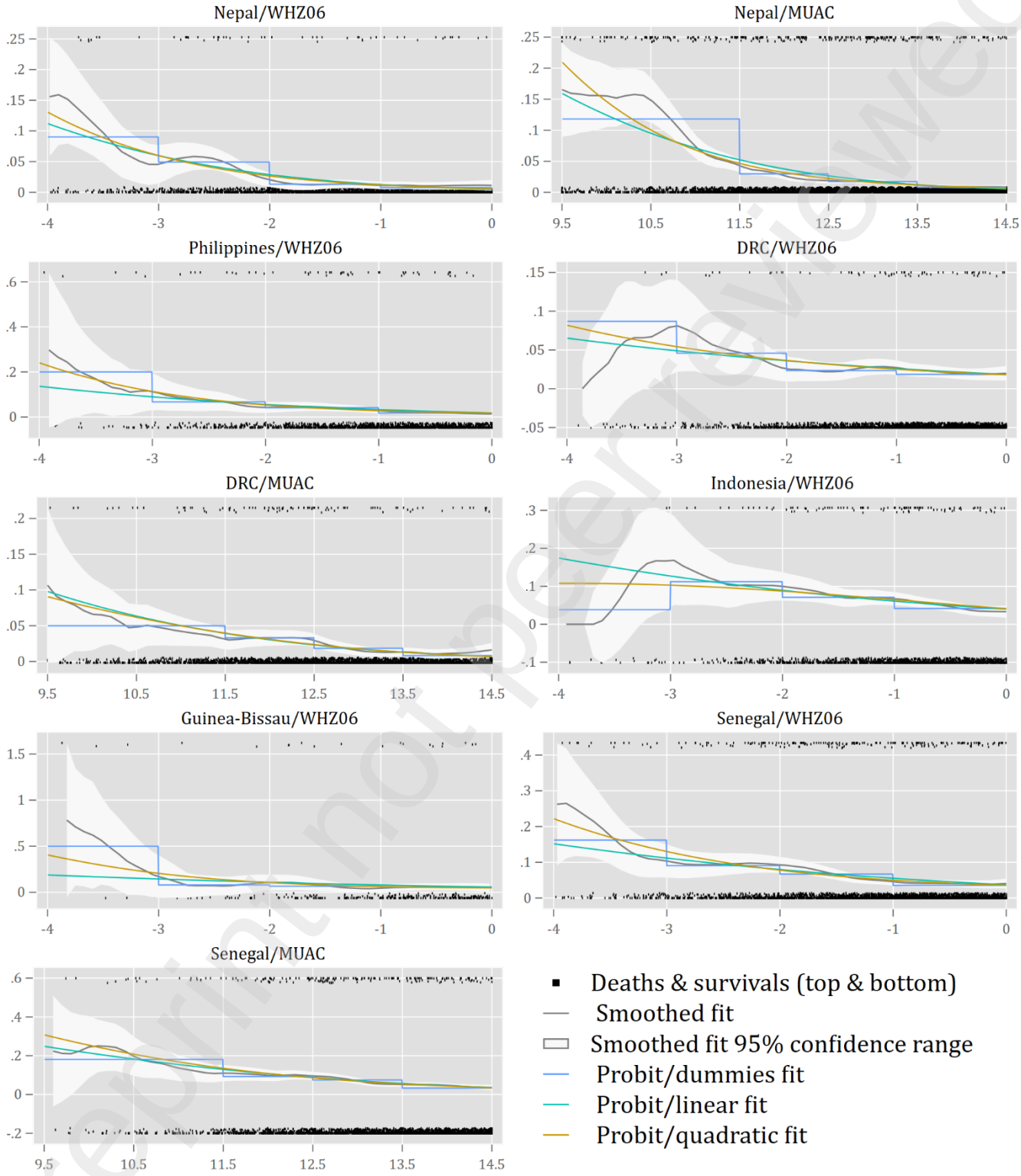
To my eye, the probit/quadratic model matches the smoothed fits *no worse* than the probit/dummies model does, even while avoiding the somewhat garish "staircase" assumption that impacts are the same within set ranges. One could also explore models with more dummies—for ranges widths of 0.5 or 0.25. But these would introduce more parameters and consume more degrees of freedom in the fitting; and the coefficients for those in sparsely inhabited categories would be estimated with great uncertainty.

I therefore prefer the quadratic model for purposes of impact estimation.

Having made that judgment, I return to the model in the previous subsection—a probit model for one-year spells, including controls—and mutate it to a quadratic specification. The structural change forces me to replace the anthropometric ranges used to this point, such as WHZ < –3, with exact values. Since the boundaries between the four standard WHZ ranges are –3, –2, and –1, it seems least arbitrary to pick the values –3.5, –2.5, –1.5, and –0.5 to stand in for the ranges. Corresponding values for MUAC (adding 14.5) are 11, 12, 13, and 14. Figure 7 shows the risks and risk ratios at those points as predicted by the fitted quadratic probit model with controls. Once again the methodology changes more than the results.

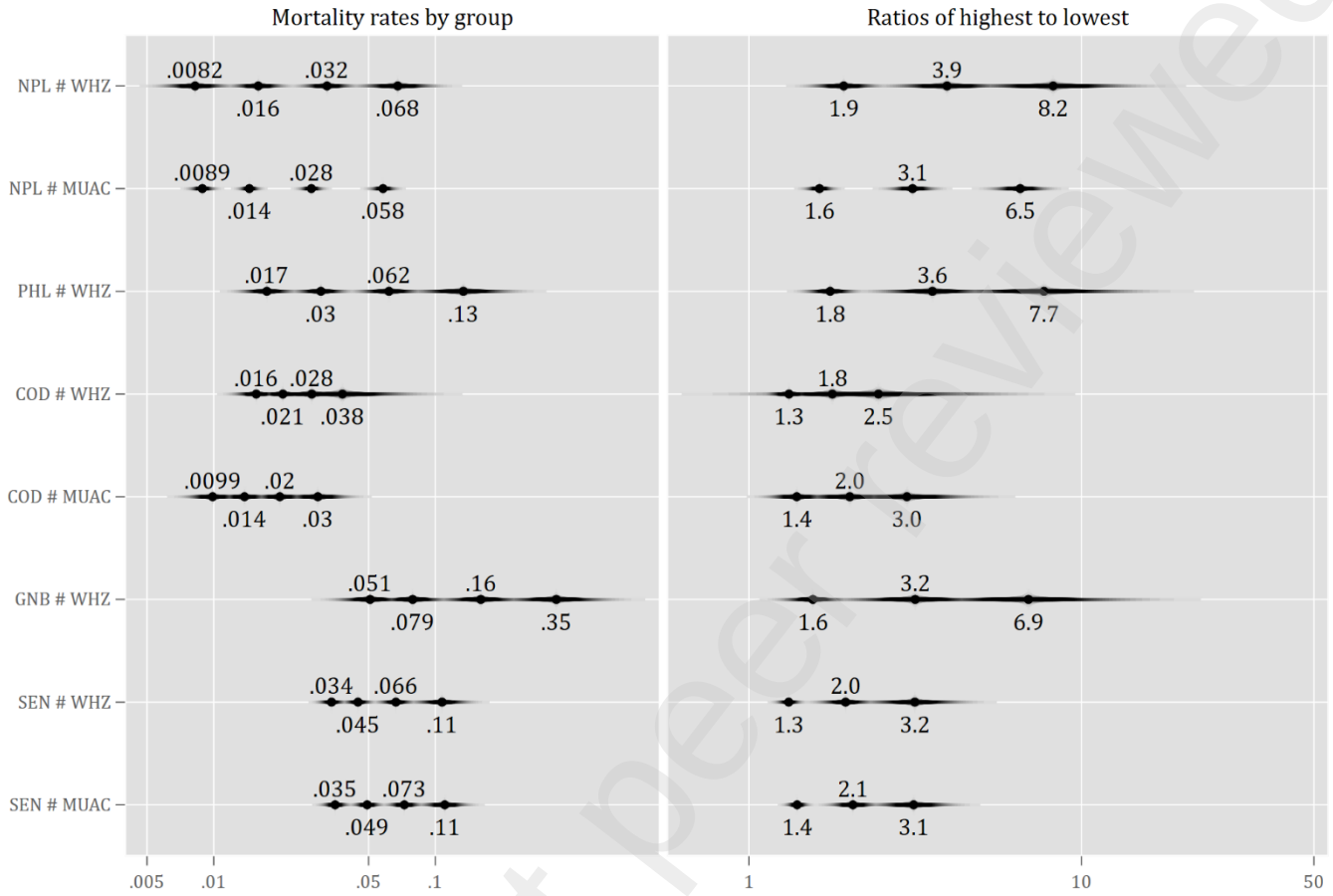
[17] The figure is inspired by graphs that Megan Higgs showed me.

Figure 6. Mortality vs. anthropometry in single-spell data sets, with smoothed fits and probit fits with dummies, linear, and quadratic specifications



Notes: Unit of observation is the one-year episode starting from a child's first measurement between age 6 and 60 months. For legibility, data for WHZ < –4 or MUAC < –10.5 are omitted. For the smoothed fits, the Epanechnikov kernel is used with a bandwidth of 0.25.

Figure 7. Predicted deaths/episode by anthropometric group, and ratios thereof to rates for healthiest group, one-year spells, probit model with Olofin et al. controls where available and anthropometric entered quadratically, with WHZ = –3.5, –2.5, –1.5, –0.5 or MUAC = 11, 12, 13, 14cm



3.7 Not incorporating timing information

The last concern addressed in this section is the disuse of information about the timing of death.

The foundational tool in survival analysis, the Kaplan-Meier survival curve, is constructed by tracking a pool of subjects over time, and in each unit of time computing the fraction of subjects who started that bit of time and made it to the end. The survival fractions are multiplied together to give cumulative survival rates over longer periods. Subtracting the moment-by-moment or cumulative survival probabilities from 1 gives the empirical hazards and the cumulative mortality rates.

Figure 8 and Figure 9 show Kaplan-Meier mortality and hazard curves for the original, multi-spell data sets, defining time by child's age. For legibility, the confidence intervals are narrowed to the 50% level. And because the hazard rates are volatile day-to-day—deaths are rare—they are smoothed, according to defaults in Stata. The vertical mortality scales are in logarithms; this way, if ratios in cumulative risk or hazard are constant, curves will move in parallel.

Cumulative mortality for the SAM groups (in black and grey) is strikingly high, exceeding 50% in some data sets. Bear in mind though that these high risks are faced only by hypothetical children who *always* qualify as SAM, whereas children typically jumped between anthropometric categories.

Some groups of hazard curves do seem to move in parallel, notably those for Senegal. Others do not, a prime example being in the Nepal/MUAC data set that is by far the largest. Tentatively, I suggest a connection between these results and the changes that occurred when moving from multi-spell to single-

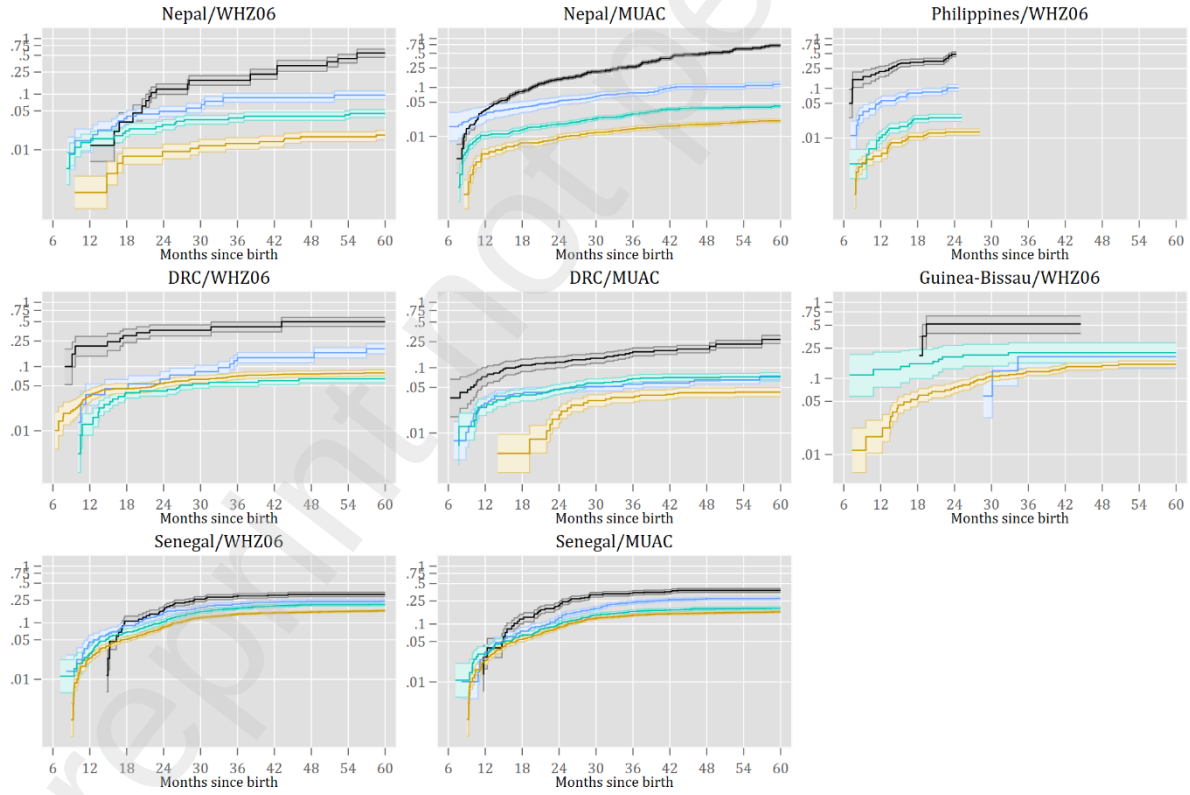
spell data (from Figure 2 to Figure 3), which I also suggested pointed to a violation of the proportional hazards assumption. In Senegal, for example, the highest-to-lowest risk ratios changed only slightly with data restructuring, from 4.7 and 5.7 for WHZ and MUAC to 4.6 and 5.5 while the Nepal numbers plunged from 28.9 and 31.8 to 11.7 and 14.8.

Despite some continuing doubt that the proportional hazards assumption holds in all the data sets, I next run Cox proportional hazards regressions, which assume that it does. The reasons: 1) perfectionism in the analysis of non-experimental data leads to nihilism; 2) the method is standard; and 3) the method is used in Olofin et al., which has influenced GiveWell's assessment of CMAM and thus is worth replicating.[18]

Cox regressions estimate hazard ratios—not absolute hazard rates, nor ratios in cumulative failure rates over stretches of time. I therefore present the Cox results in the format of the right panes of earlier figures. See Figure 10. I include the Olofin controls and, as reported in the top and bottom halves of the figure, run regressions on the both the multispell and single-spell data. The former constitute my closest replication of the WHZ regressions in Olofin et al. (2013, Table S2(C)). Original and replication are recognizably similar: for example, the highest-to-lowest hazard ratio for the Philippines is 39.0 there and 33.3 here—in both cases the largest of all.

The Cox hazard ratio estimates are remarkably similar to the earlier cumulative mortality estimates: compare the upper half of the new figure with the right pane of Figure 2 and lower half with the right of Figure 3—or, even better, Figure 5, which also comes from regressions with controls. The match is especially good in the latter case.

Figure 8. Kaplan-Meier failure curves by country, anthropometric indicator, and four-way grouping, with 50% confidence intervals



[18] "Our starting point to estimate mortality among untreated malnourished children comes from Olofin et al. 2013." givewell.org/international/technical/programs/combined-protocol-community-management-acute-malnutrition.

Figure 9. Smoothed hazard rates per month by country, anthropometric indicator, and four-way grouping, with 50% confidence intervals

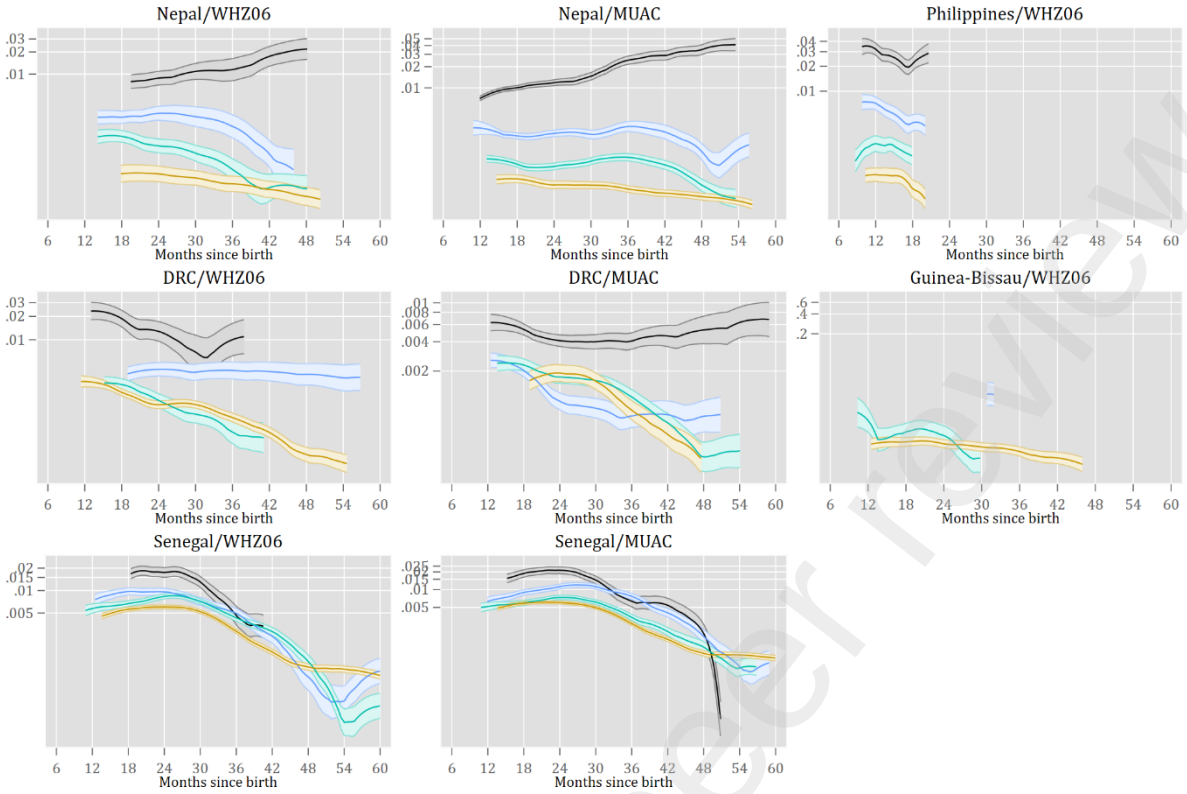
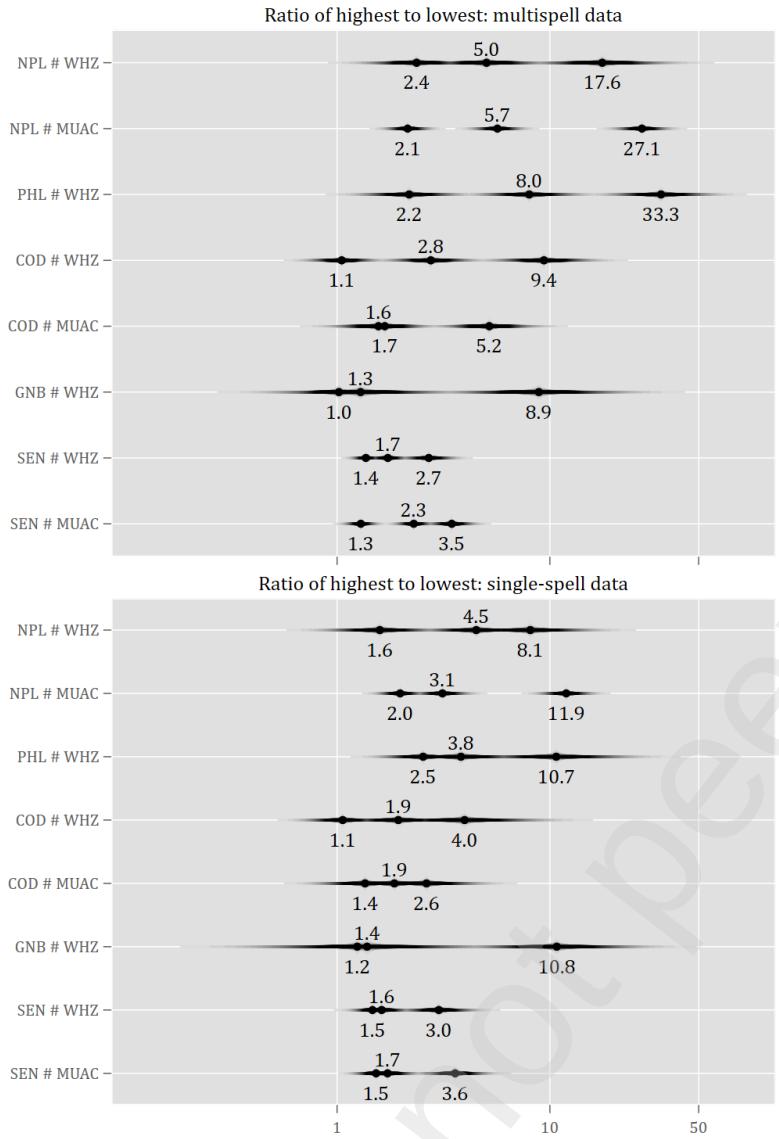


Figure 10. Cox hazard ratio estimates on multispell and single-spell data sets, including Olofin controls where available



3.8 Summary

This section began by computing simple mortality rates and ratios thereof. Motivated by concerns about how those simple statistics could mislead if extrapolated directly to treatment settings, it introduced elaborations: moving to one-year spells, adjusting for age by restricting the sample, moving to a probit model in order to introduce controls, exploring alternatives to dummies for representing anthropometry, applying non-parametric and parametric methods from survival analysis.

Only one of the changes made much difference: moving to one-year spells. This is good to know. But it also poses a mystery. By arbitrarily modifying the data sets, we can raise and lower the risks and risk ratios for various anthropometric groups. We could, for example, use 6- or 18-month spells instead of 12-month ones, or stay with multi-spell data but merge successive pairs of episodes. What then does any particular set of results mean?

It is not clear what sort of real-world processes would generating data that would behave in this way. That is, we lack theory. And without a theory that roughly explains the evidence, it is harder to decide how to extrapolate to modern treatment contexts.

In the next major section, I will describe a theory, one that, unlike the Cox model, represents the data generating process in way that gives an explicit role to the passage of time. If the model is exactly correct, then estimates based on calibrating it to real data will be nearly impervious to changes in episode length. Inevitably, the model is simplistic and wrong, but it does prove to be more stable than the Cox model, or simple averages for that matter, while still allowing incorporation of age and other controls and quadratic dependence on anthropometry.

4 The problem of unobserved heterogeneity

It is tempting to posit, as the Cox model does, that risk ratios are constant over time. Sometimes that makes sense: if for reasons of geology, the probability of being hit by lighting is twice as great in Oklahoma as in Vermont, that is probably true over minutes and millennia.

But in general there is a good reason to expect that risks and risk ratios change: *unobserved heterogeneity.* Within an experimental treatment group, for example, a new drug might work better for some people than others, perhaps because of genetic differences. Then, to put it crudely, after some time has passed, the people who are going to die will have already died. The risk of death in what remains of the treatment group will fall. As a result, the drug trial will return different hazard and cumulative risk ratios if run for a month than if run a year.

It is perhaps intuitive to conceive of two forms of unobserved heterogeneity. The first, *frailty*, was just exemplified. Factors distinct from those measured by researchers influence outcomes; subjects who appear the same in the data in fact differ in crucial ways; among same-seeming subjects, those put at most risk by these unobserved factors exit the risk pool first.

The other form of unobserved heterogeneity is *aging of measurement*. It stands to reason that the risk associated with a red-zone reading on a MUAC measurement taken today speaks more to a child's risk of death now than one taken a year ago.

Though the two forms of unobserved heterogeneity look different, deeper down they are the same. Old MUAC measurements are associated with lower risk ratios because after those old measurements are taken, children's anthropometric statuses evolve in different directions, for reasons unobserved by the researcher. That is frailty by another name.

We have already seen signs of measurement aging. Converting data sets with typically 2–6-month episodes, thus typical time since measurement of 1–3 months, to data sets with 12-month episodes significantly reduced risk and hazard ratios, whether working with simple averages or probit or Cox models with controls.

Here I present further evidence of such effects.

The first demonstration is motivated by the idea that if the proportional hazards assumption is indeed violated in real data, perhaps proportional hazards regression can be salvaged as a practical tool if we apply it to narrow time windows, within which the hazard ratio will change less. In this spirit, I run a Cox regression that models the hazard of mortality only in the first month after a child's height and weight were last measured. Then I shift the window to the second month, and so on.[19] I pool the data from the five countries with timing information and stratify the model by country.

[19] Regressions are stratified by country. Standard errors are clustered by individual. Spells leading to failure before the start of a window are dropped. Spells leading to failure after the end of a window are recoded as non–failures. Because of a tendency to degeneracy in these restricted regressions, all controls are dropped to aid convergence.

Figure 11 shows the resulting estimates and confidence bands for the hazard ratios for WHZ < -3 (SAM) vs. WHZ > -1. In the first month after measurement, the hazard is 35 times higher in the SAM group. The ratio falls as the window shifts to months 2–5.

For a second demonstration, I focus on the data from the Philippines study, the one in which children were measured at the highest frequency: every two months for up to two years. I run a Cox regression on the full sample. Then I drop every other observation for each child, to simulate them being measured every four months instead of every two, and rerun the regression. I continue the pattern by retaining only every third observation, and so on. Again, the hazard ratios start high and fall as the time span lengthens.[20] (See Figure 12.)

Finally, I extract and rearrange some of the multispell Cox results that were reported in the top half of Figure 10. The novelty is to plot the hazard ratios as a function of average episode length (for survivors). For legibility, I take the hazard ratios for the SAM category only, and focus on WHZ. (The WHZ and MUAC results would collide since they would have the same x coordinate and similar y coordinates.) See Figure 13.[21] Again, where children are followed up on more quickly, hazard ratios appear higher.

All of these results suggests that frequency of follow-up is a major determinant of apparent cross-country differences in hazard ratios.[22] It appears that much of the cross-country variation within published meta-analyses is an unrecognized artifact of aging of measurement.

[20] Regressions as described in note 19, except restricted to the Philippines. The hazard ratio of 126 differs from the corresponding 39 in Olofin et al. (2013), Table S2(C) for reasons I cannot explain; the corresponding author of Olofin et al. did not share the code for that study. However, the replication squares with the original more than it may appear in that the natural domain for Cox hazard ratios is logarithms, and in logarithms 39 and 126 are less distant. Moreover, among the Olofin et al. country–specific weight–for–height results, that for the Philippines is distinctly highest.

[21] Average is of spell lengths not ending in death.

[22] The average spell length in each country reflects the precise timing of visits, of any deaths between visits, and, in some cases, of an endline mortality survey.

Figure 11. Results from five-country Cox regression when shifting window within which failure is modeled

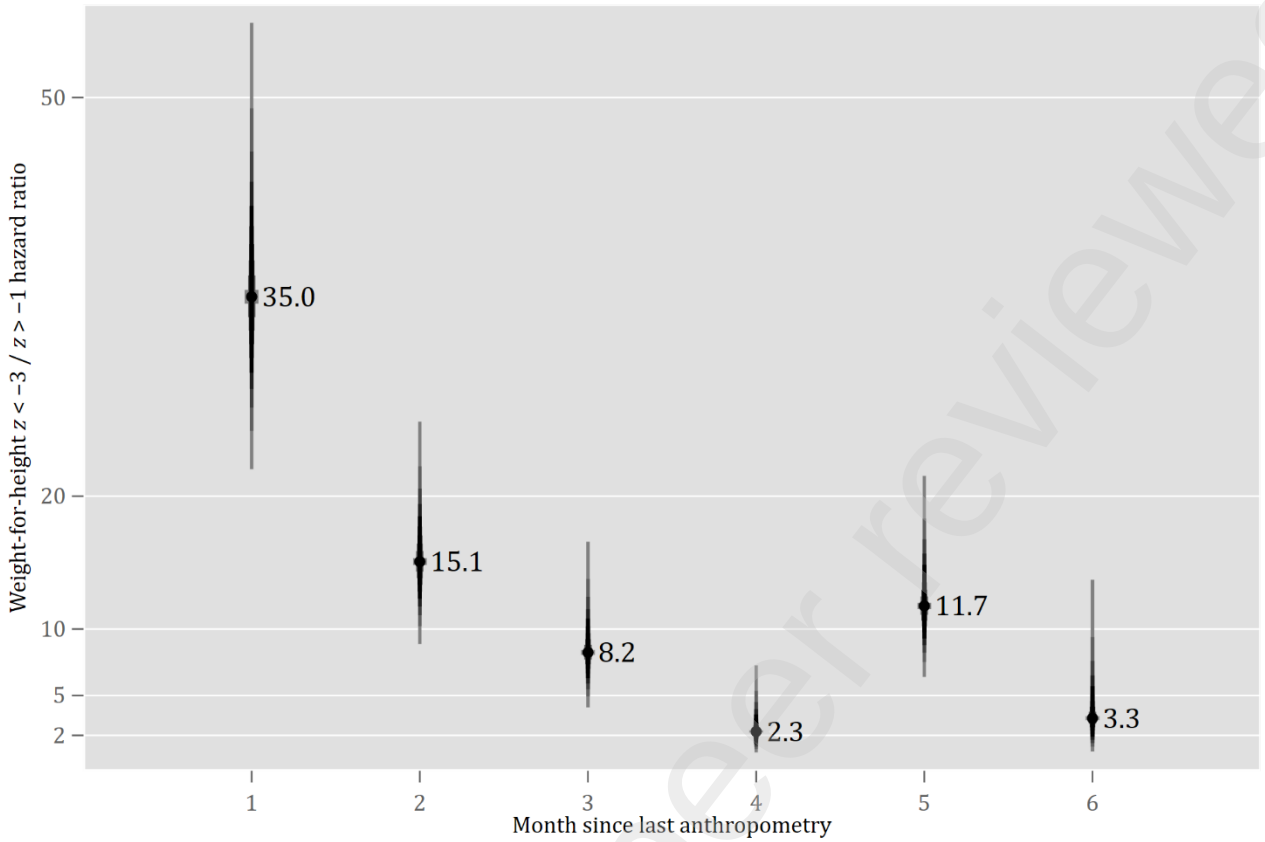


Figure 12. Results from Philippines Cox regression retaining only every n^{th} bimonthly observation for each child, $n = 1,2, \ldots$

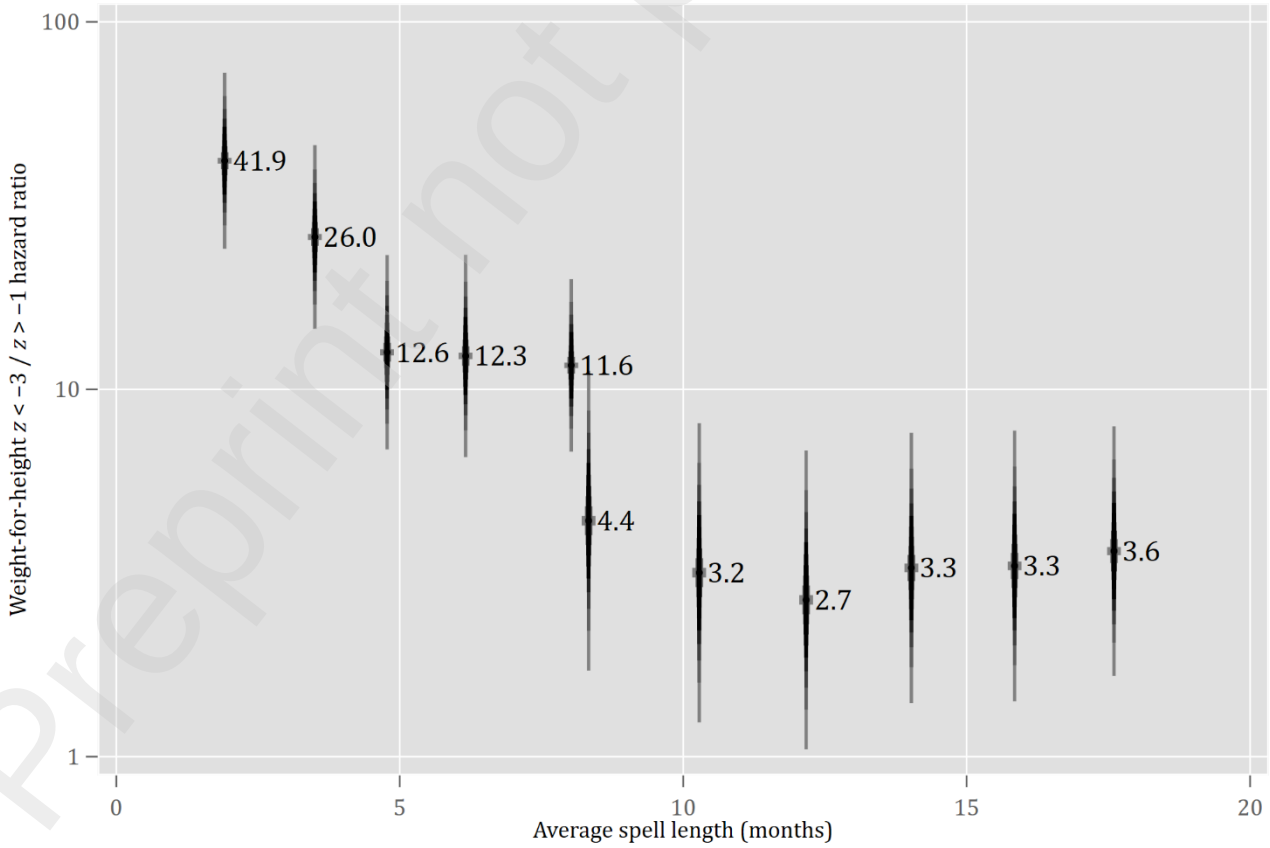
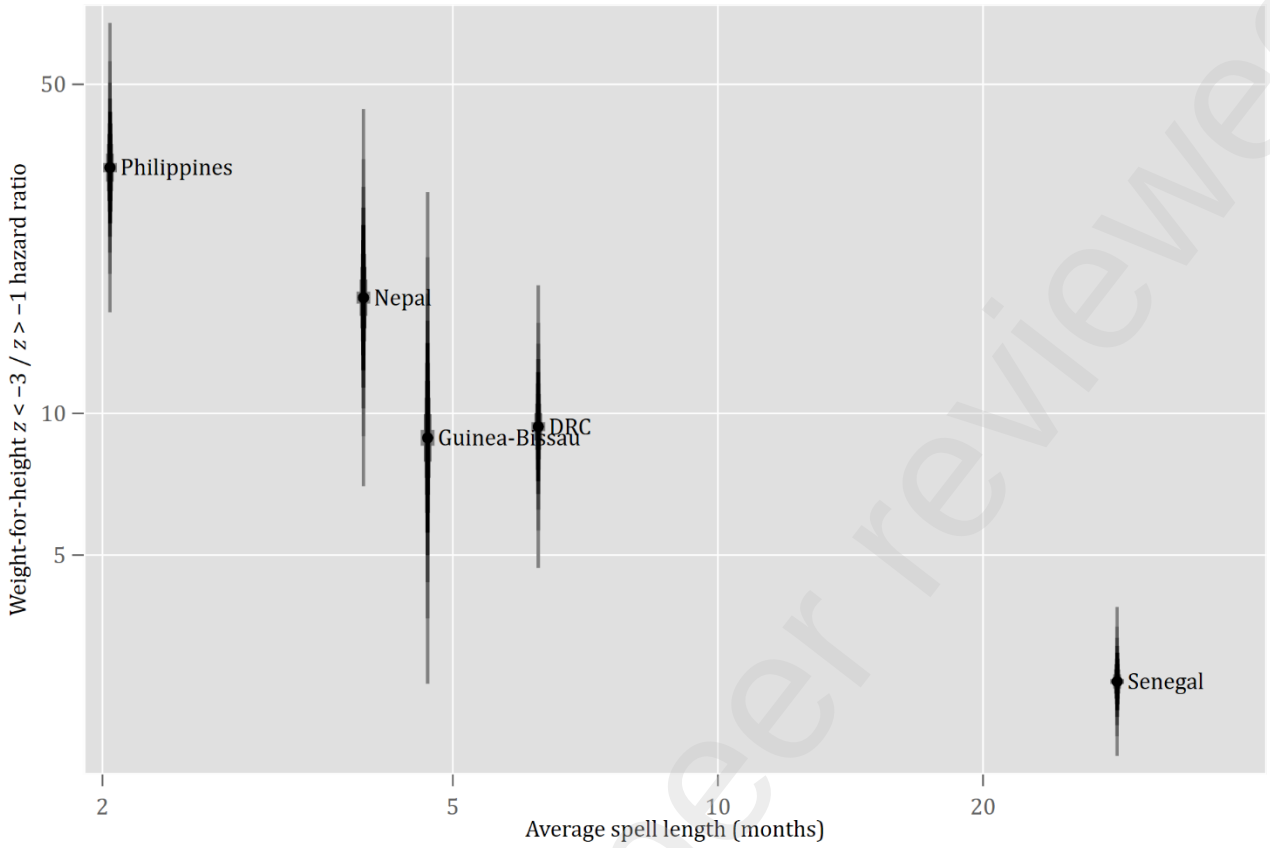


Figure 13. Cox regression results versus average spell length in five countries with data



5 Mixture Inverse Gaussian (MIG) modeling

This section introduces and applies an alternative to the Cox proportional hazards model, one that posits an explicit microtheory for the evolution of health over time. Use of the model has three potential benefits:

- To the extent that the model captures the patterns in the data, it will be robust to aging of measurements, producing more-comparable results from data sets with various follow-up frequencies.
- Once calibrated to data, the model can generate predictions of cumulative risks as well as hazards for any time point since measurement, the first being closer to the outcome of interest in assessing CMAM. And it allows predictions of absolute risks, not just ratios between risks.
- Because it contains a mathematical theory for the evolution of health, it could provide a firmer basis for extrapolation of results from historical settings to prospective treatment settings. A weekly feeding program, for example, could be represented in the model as constant upward pressure on the health level, and the consequences for mortality could be simulated from there. This is not possible with models such as the probit and Cox that lack such foundation.

I chose the model after recognizing the measurement aging effects in a graph like Figure 13. The pervasiveness of unobserved heterogeneity and the consequent threat to the proportional hazards model are leitmotifs in the Aalen, Borgan, and Gjessing (2008) textbook I was reading. A quick search for practical alternatives—in this book and in the Stata manuals—surfaced only one that appeared to meet the following criteria:

- *Explicit representation of the role of time.* This seems needed for robustness to cross-study differences in follow-up frequency.

- *Tractable formulas for the survival probability and failure time distribution.* Most stochastic models one can write down do not have closed-form solutions, and can only be fit through a computationally intensive and algorithmically complex use of simulation (Hurn, Jeisman, and Lindsay 2007). Random-effects shared frailty models are common, but these require the practitioner to identify and observe the dimension of shared frailty, such as the school or province. In the case at hand, I see no compelling candidates—especially in the DRC and Senegal data sets, which include essentially no covariates.

The model meeting these criteria is developed in Aalen (1994) and reintroduced and applied in Aalen, Borgan, and Gjessing (2008), chapter 10. The drive for tractability shapes the model, giving it a straitened (and straightened) character.

5.1 The MIG model

In the model, the state of a given individual at a given time is characterized by the abstract variable "health." We do not observe "health" except in that if it falls to zero, the individual dies. In the simplest version, everyone's health starts at the same level. From there, each individual's health tends to rise or fall at a constant rate during an episode. Yet two components of randomness enter. First, the slopes of the individual health trajectories are not all the same; rather they are randomly assigned from a normal distribution defined by some average and variance. This variance constitutes the unobserved heterogeneity. Second, all individuals experience moment-to-moment random changes in their health levels, which accumulate—what is called Brownian noise.

This data generating process may be termed "Brownian motion with random drift." The probability distribution it produces for time of death is the Mixture Inverse Gaussian (MIG). So I call it the MIG model.

Formally, in Brownian motion with random drift, individual i's health, $x_i(t)$, starts at c, at time 0:

$$x_i(0) = c$$

Health evolves according to the stochastic differential equation

$$\begin{array}{ll} dx_i = -\mu_i dt + dB_i & \text{if } x_i > 0 \\ dx_i = 0 & \text{if } x_i = 0 \end{array}$$

μ_i is the individual-specific slope (drift) that characterizes the individual's trajectory. The minus sign on μ_i establishes the convention that higher values of the parameter reduce health, i.e., increase the probability of death. dt represents an infinitesimal increment of time. dB_i represents an infinitesimal increment of Brownian noise, which is a fractal random walk and has cumulative variance of σ^2 per unit time. Because health, x_i, is not observed unless it hits 0, the scaling of the x variable is arbitrary and immaterial. We normalize scale by setting $\sigma^2 = 1$.

The average slope of each individual's health trajectory, μ_i, is drawn from an another, independent normal distribution, with mean μ and variance τ^2:

$$\mu_i \sim \mathcal{N}(\mu, \tau^2)$$

μ_i is constant for each individual in each episode, but can change arbitrarily when a new episode starts.

The boundary at $x_i = 0$ is "absorbing": any path reaching the boundary remains there forever, since death is irreversible.

Three parameters thus shape the stochastic process: c, μ, and τ^2. In terms of those parameters, Aalen (1994, p. 240) derives formulas for the probability of survival and the timing of death—in mathematical terms, the probability that $x(t) > 0$ given t, and the probability density over t that x first hits 0 at time t.

To be precise, the probability distribution for $x = x_i(t)$ when $x > 0$ is

$$\psi(t) = \left(1 - e^{-\frac{2cx}{\sigma^2 t}}\right) \phi(x; c - \mu t, \sigma^2 t)$$

where $\sigma^2 = 1$ and $\phi(\cdot;\cdot,\cdot)$ is the normal density with indicated mean and variance. Since x is not observed, the above equation is not feasible and is not used in model fitting. However, its integral in x over $(0, \infty)$ is the survival probability at time t, which does correspond to an observable event. The probability works out to

$$S(t) = \Phi(c - \mu t; \tau^2 t^2 + \sigma^2 t) - e^{2\frac{c}{\sigma^2}\left(\frac{c}{\sigma^2}\tau^2 + \mu\right)} \Phi\left(-\mu t - c - 2\frac{c}{\sigma^2}\tau^2 t; \tau^2 t^2 + \sigma^2 t\right)$$

where $\Phi(\cdot;\cdot)$ is the cumulative normal density with indicated variance. And the negative of the time derivative of that is the temporal distribution of death:

$$f(t) = \frac{c}{t} \phi(c - \mu t; \tau^2 t^2 + \sigma^2 t)$$

The latter two equations are used in model fitting, the first for survivors and the second for non-survivors.

By definition, cumulative mortality and instantaneous hazard rates are then

$$F(t) = 1 - S(t) \tag{1}$$

$$h(t) = \frac{f(t)}{S(t)} \tag{2}$$

The three model parameters can vary by individual, through dependence on observables such as WHZ and age at measurement. For example, I model with

$$c = \boldsymbol{\beta}_c' \mathbf{z}$$

$$\mu = \boldsymbol{\beta}_\mu' \mathbf{x}$$

τ^2 is the same for all individuals

$\boldsymbol{\beta}_c$ and $\boldsymbol{\beta}_\mu$ are parameter vectors for estimation, along with τ^2. Recall that c is the starting point and μ the average rate of health progression or regression from there. The vector $\mathbf{z}$ contains, in addition to a constant term, one or more variables that depend on WHZ or MUAC, such as those variables and their squares. $\mathbf{x}$ includes a constant, as well as gender, age at last measurement, and potentially additional controls defined by Olofin et al. (see Table 3 below).[23] $\mathbf{x}$ also contains the anthropometry in $\mathbf{z}$ in order to allow for regression to the mean, for example, because of measurement error.

The overall structure—with initial health depending on a diagnostic, subsequent progression depending on a larger set of potential ongoing influences, and the slope variance τ^2 held the same for all individuals—largely follows the pattern in the application to oropharyngeal cancer in Aalen, Borgan, and Gjessing (2008), §10.3.8.

The model is fit with Maximum Likelihood using a Stata program I wrote. The unit of observation is the child-episode. As noted, a child's health trajectory μ_i can vary from one episode to the next; but it will vary less to the extent that the model fitting associates μ with slow-changing variables such as ethnicity and household size.

[23] I initially fit with $\ln c = \boldsymbol{\beta}_c' \mathbf{z}$ to guarantee that trial values of initial health are positive, then switch to the un-logged model and refine.

Notably, this model departs from many survival models in measuring time from the beginning of a spell—the time of last measurement—rather than from a universal anchor such as time of birth or marriage. If two individuals were last measured two months ago, but at different ages, their data plays the same role in the MIG model: contributing to the assessment of risk two months after measurement. In typical applications of Cox and Kaplan-Meier methods, such as in section 3.7, the experiences of the two individuals would contribute to risk assessment at their respective ages. By the same token, the MIG model does not exploit the multispell nature of the data at hand. Each episode of measurement and follow-up on a given child is treated in parallel, with no adjustment or exploitation of the fact that the episodes are for the same child.

Those comments on the structure of the model are subject to two caveats. First, as noted, age *does* enter the MIG model applied here in another way: it appears in the μ equation as a determinant of individual health trajectory. Second, standard errors can be clustered by subject, and here are.

5.2 Illustration

A sequence of graphs helps make the ideas in the MIG model more concrete. Figure 14 shows 99 sample paths from a Brownian motion with random drift, run for a simulated year. For this illustration, all the paths start at $c = 1.82$, a value that arose when I calibrated a version of the MIG model to the data from Guinea-Bissau; the value is abstract, but in context stands for children who have been diagnosed with WHZ-SAM. From the shared starting point, the paths diverge along trajectories that indeed resemble straight lines, with a bit of wandering noise. The average slope is set to be positive, at $\mu = 0.86$, so that most individuals improve with time. But the standard deviation of the slopes is $\tau = 0.57$, large enough to produce a minority of paths that head downward, some to death.[24] Notably, most of the deaths happen within the first few months. Thereafter mortality in the surviving population falls, because most of the survivors have positive trajectories.

Figure 15 shows the distribution that arises from running 10,000 paths instead of 99. Shading boundaries mark the 5th, 10th, etc., percentiles. The dark line shows the median. The lowest band plotted, for the 5th to 10th-percentiles, heads to death, as does part of the next band.

Figure 16 surfaces this aspect, showing the cumulative mortality rate. The black curve indicates the result for the 10,000-path simulation while the light blue shows the prediction from theory (equation (1) above). Just as in the Kaplan-Meier curves in Figure 8, cumulative mortality tends to plateau.

Next, I posit the existence of a second group of individuals, which differs from the first only in starting at a health level corresponding to children with WHZ > -1. Figure 17 augments the previous graph with corresponding curves for this healthier group.

Next, I take the ratios in cumulative mortality across the two groups, a measure of the elevation of risk from starting at a worse health level. (Figure 18.) It turns out that the mortality ratio is extremely high in the first few days and declines from there toward a limit of about 1.31.

The final two graphs are like the last two except that they show hazard rather than cumulative risk. Simulation again matches theory (equation (2)). The hazards start low, but quickly rise as it becomes more common for ill-fated paths in both groups to hit zero. Eventually the trend reverses: most of those downward paths do reach death, and the hazard among survivors falls. (See Figure 19.) This quick rise and fall, it should be said, is not discernible in the empirical hazard curves in Figure 9. That mismatch may indicate unrealism in the model. It may also be caused by having samples too small to finely measure

[24] The variance of the Brownian noise is 1 unit per month. This normalization is imposed because health is unobserved unless it hits zero, so the scale of the health variable is immaterial (Aalen 1994, p. 240).

gyrating hazard curves: even the simulated hazard curves depart noticeably from theory despite a sample size of 10,000.

Finally, the *hazard ratio* across the two groups, far from being constant, starts high and falls monotonically (Figure 20). That does cohere with the empirical demonstrations in the previous section.

Figure 14. 99 sample paths of Brownian noise with random drift and absorbing boundary at 0

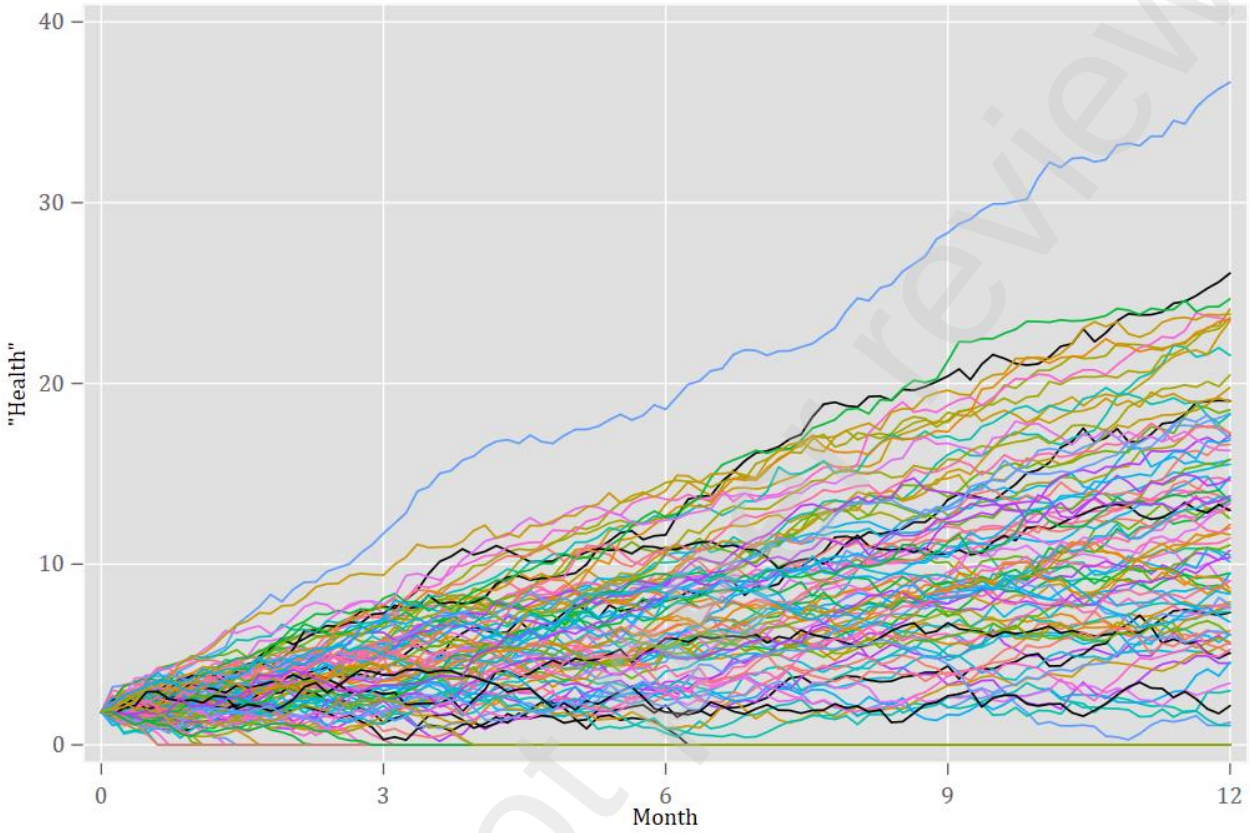


Figure 15. Distributions of 10,000 sample paths of Brownian noise with random drift and absorbing boundary at 0

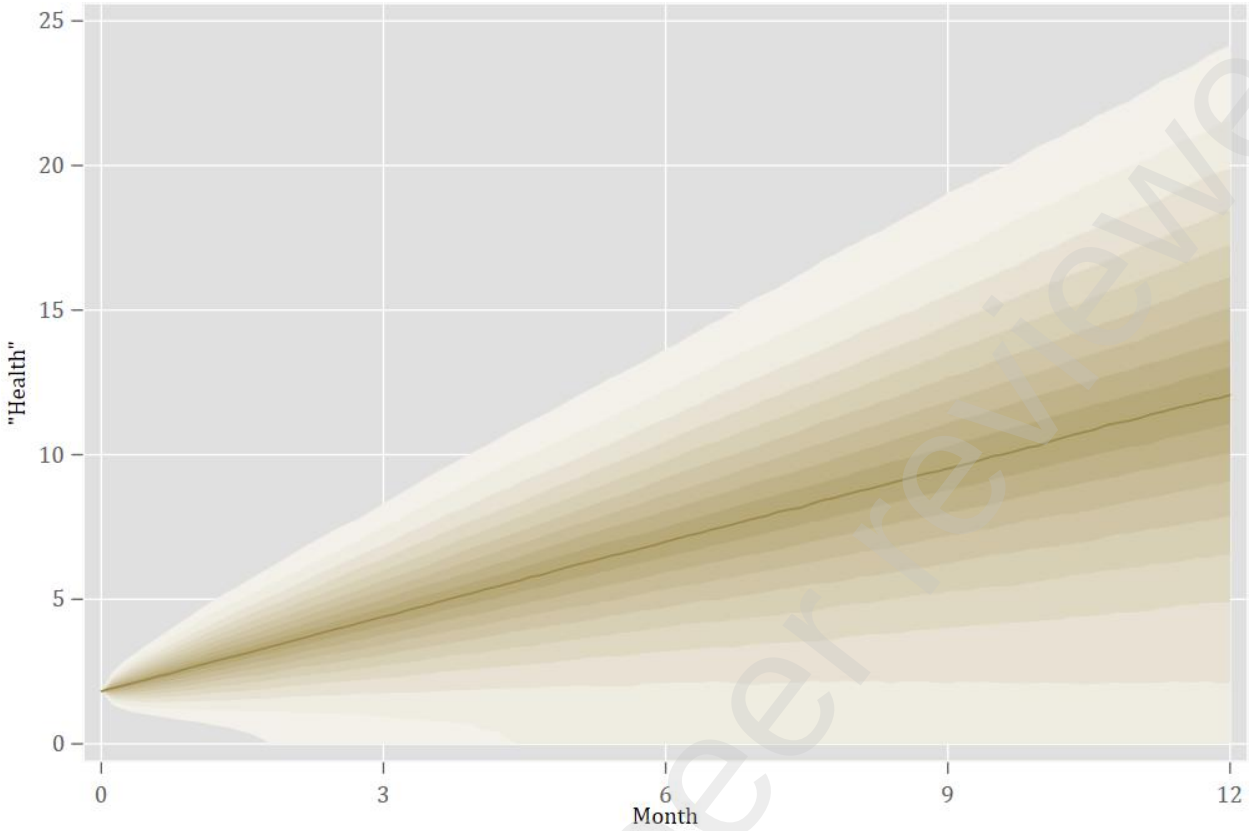


Figure 16. Simulated and theoretical mortality in cohort with common starting point

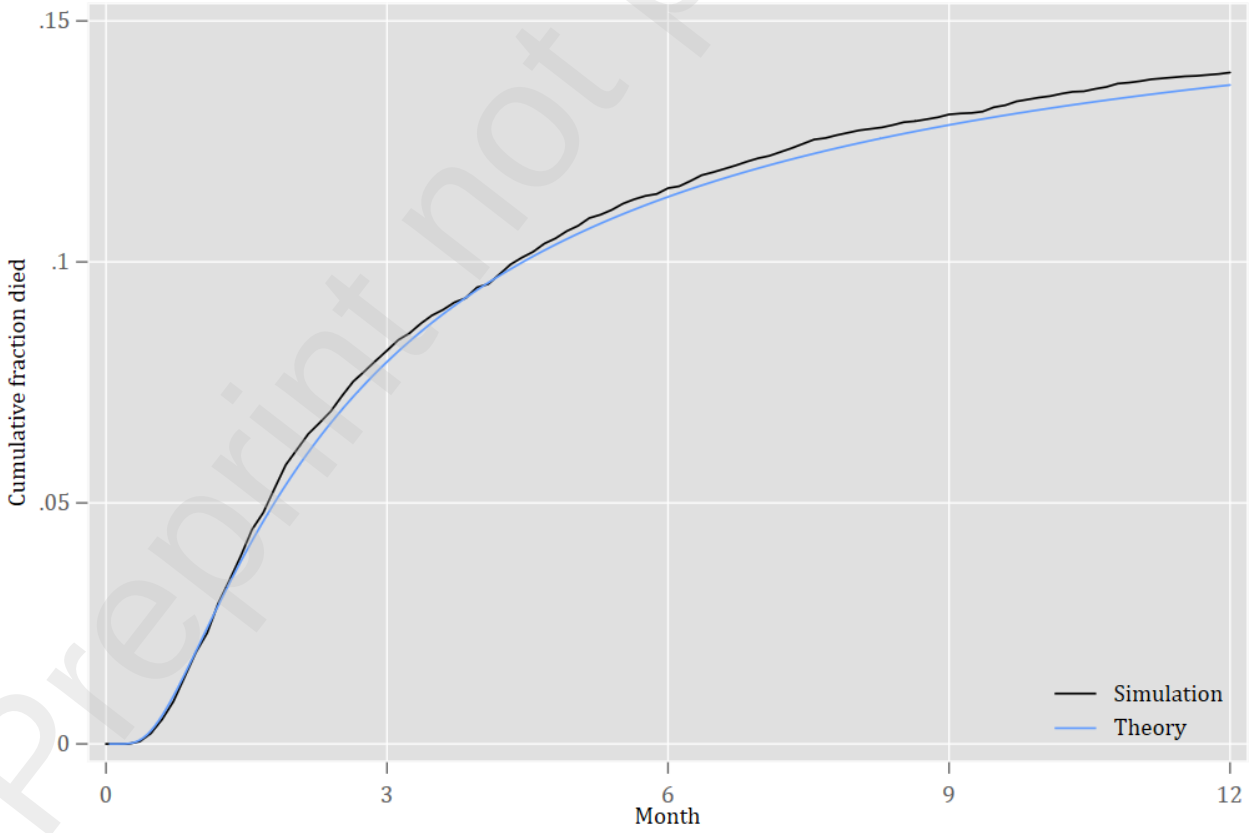


Figure 17. Simulated and theoretical mortality in cohorts with two different starting points

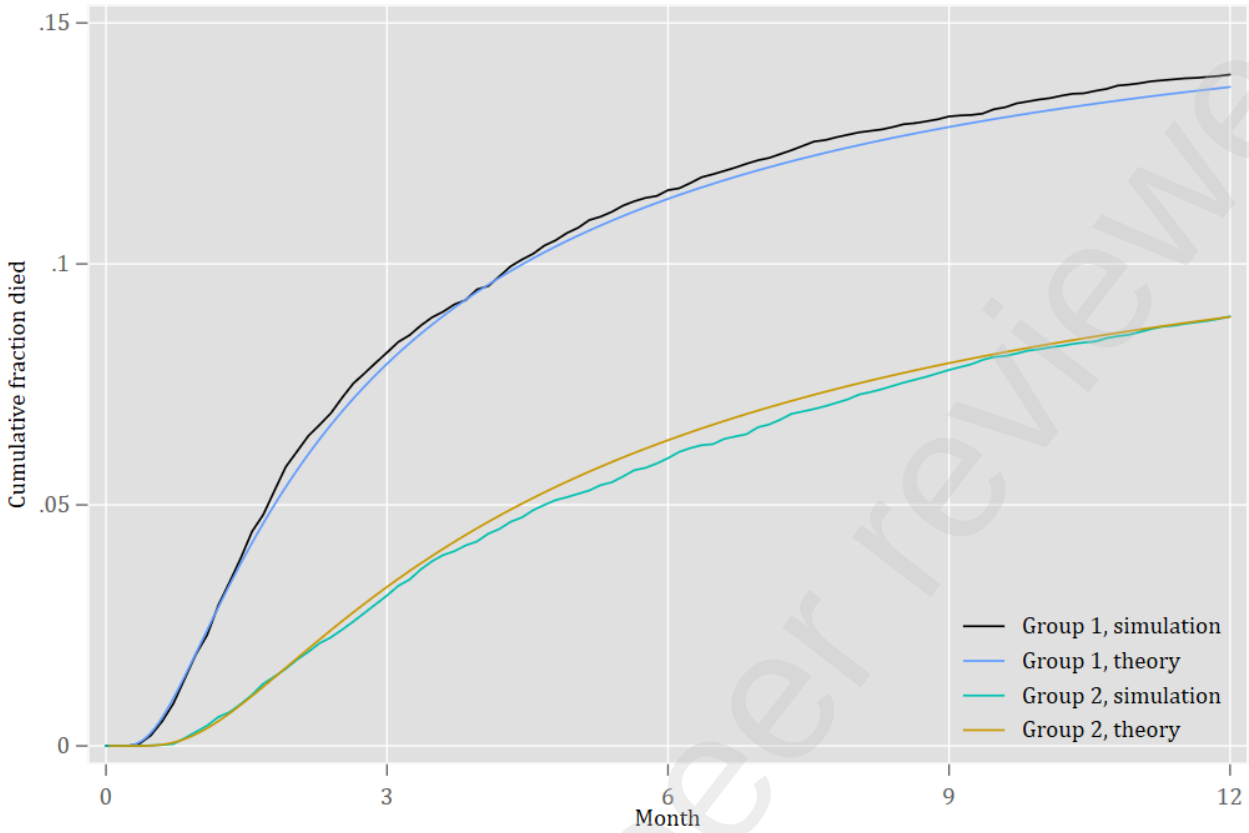


Figure 18. Simulated and theoretical cumulative mortality ratio between cohorts with two different starting points

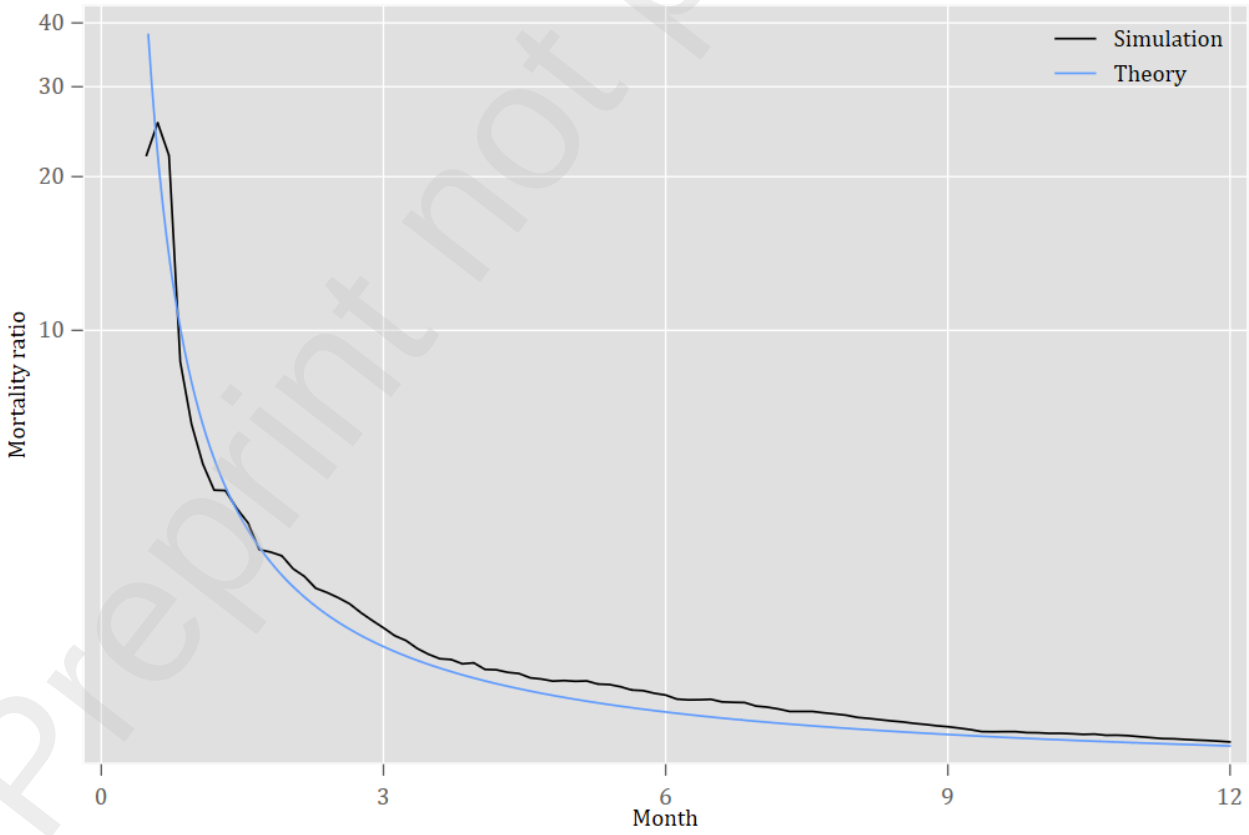


Figure 19. Simulated and theoretical hazard rates in cohorts with two different starting points

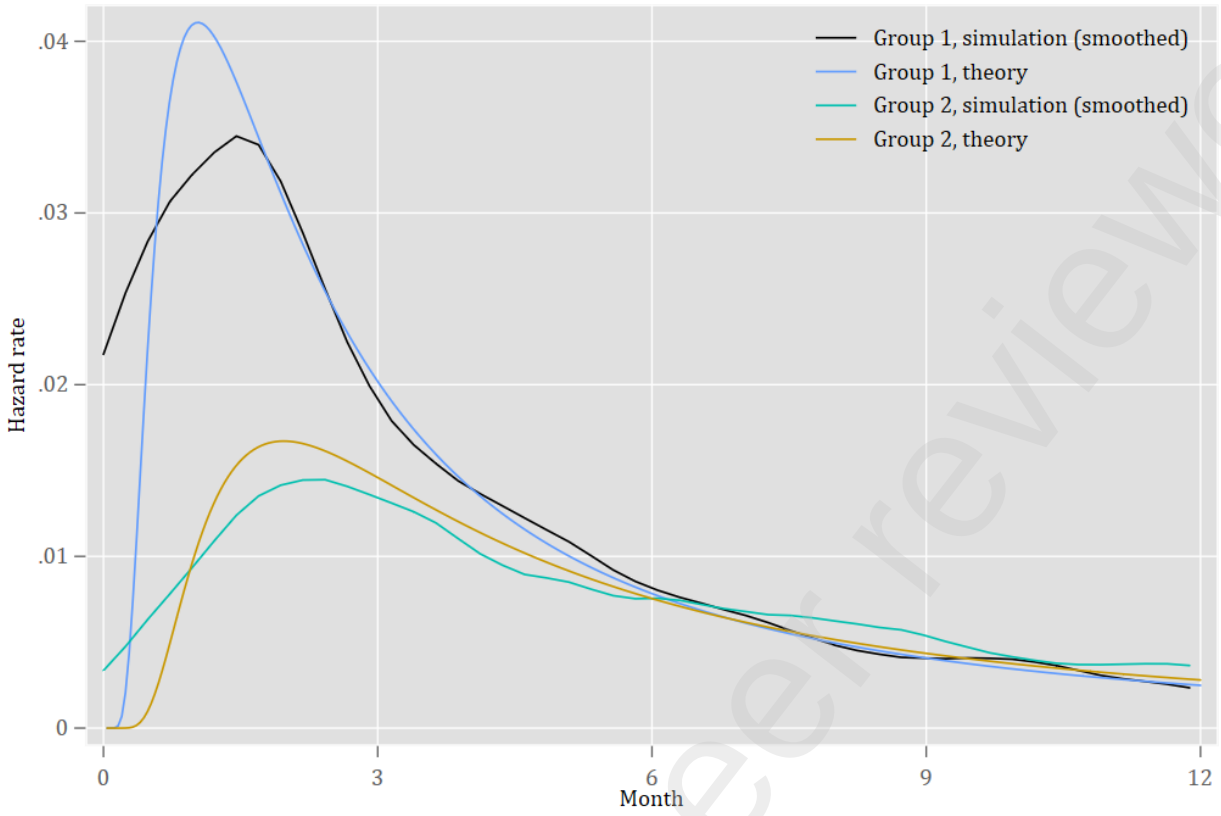
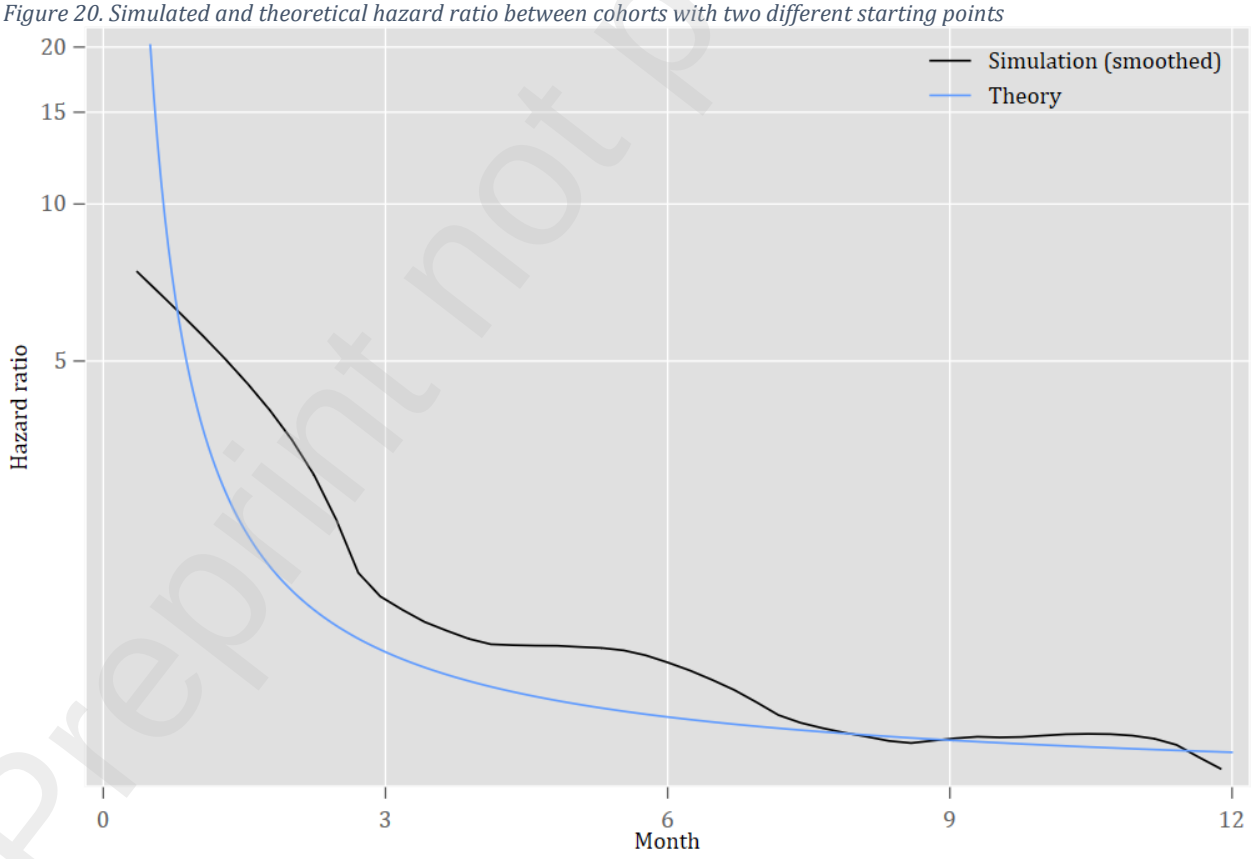


Figure 20. Simulated and theoretical hazard ratio between cohorts with two different starting points



5.3 Testing

In section 3.6, and especially Figure 6, I challenged a stripped-down probit model, one whose only regressors were variables to represent WHZ or MUAC, to fit basic patterns in the data. The fits were good enough to justify some confidence in the validity of the model. Here, I do something similar for the MIG model.

The probit model was tasked with predicting *survival rates* as a function of anthropometric status. In survival modeling proper, the analogous ground-truthing challenge is to match *survival curves*—capturing not only whether death occurs but also when. I will fit the MIG model to multispell data set since those contain the most timing and anthropometric information, and ideally the estimator is robust to the diversity in spell lengths. Since the model's clock starts with last measurement rather than birth, the benchmark survival curves will be defined in the same way. The MIG model that is tested on them includes no controls, not even for age. It only includes the familiar sets of four WHZ or MUAC dummies, in the c and μ equations.[25]

The test results are gathered in Figure 21. Once again, mortality is graphed rather than its complement, survival. The four jagged lines in each plot, bracketed by 50% confidence bands, are the purely empirical mortality curves. The smooth curves of the same color are predictions based on the best MIG fits. The time span chosen for each plot, such as two months for the Philippines, reflects the typical follow-up frequency in each data set.

The simplistic model fits the data surprisingly well. There is perhaps a tendency to overestimate mortality in the healthiest group (light orange), which could downward-bias risk ratios expressed with respect to that group.

Next, I return to the earlier narrative: is the MIG model indeed robust to episode lengthening? Just as with the Cox model in Figure 10, to produce Figure 22 I run the MIG on each of the data sets in the multispell and single-spell configurations and plot the predicted risk ratios. The estimator appears more stable than the Cox, though not perfectly stable. For example, where the SAM risk ratios in the Nepal data fell from 17.6 to 8.1 and from 27.1 to 11.9 for WHZ and MUAC respectively, the MIG risk ratio estimates move from 11.3 to 5.9 and 5.7 to 5.9.

Overall, the MIG model looks stable and accurate enough that its full results, like the probit model's, deserve some credence.

[25] c is modeled in logarithms in order to guarantee positivity.

Figure 21. Kaplan-Meier survival curves in time since measurement, by country, anthropometric indicator, and four-way grouping, with 50% confidence intervals and MIG model fits

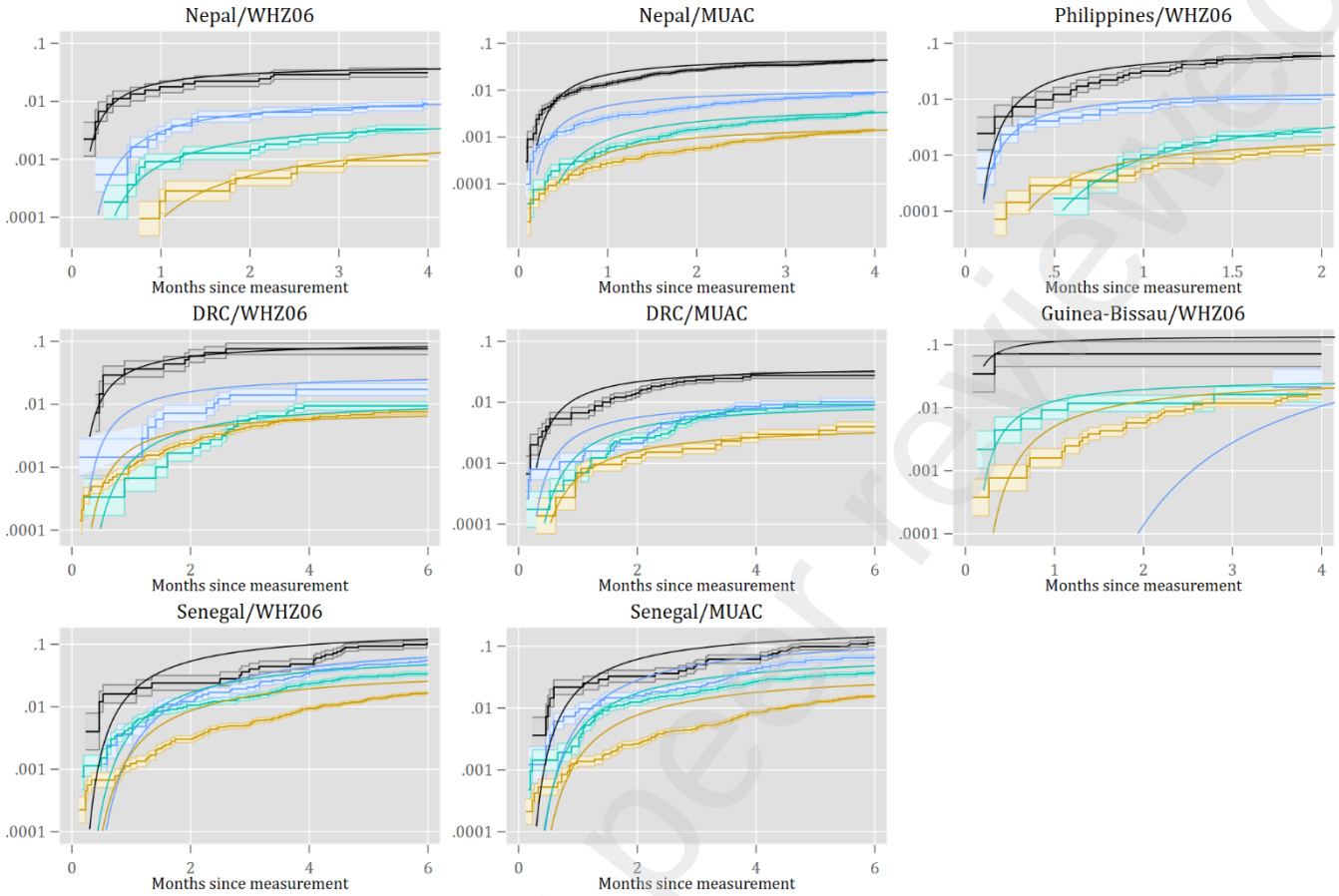
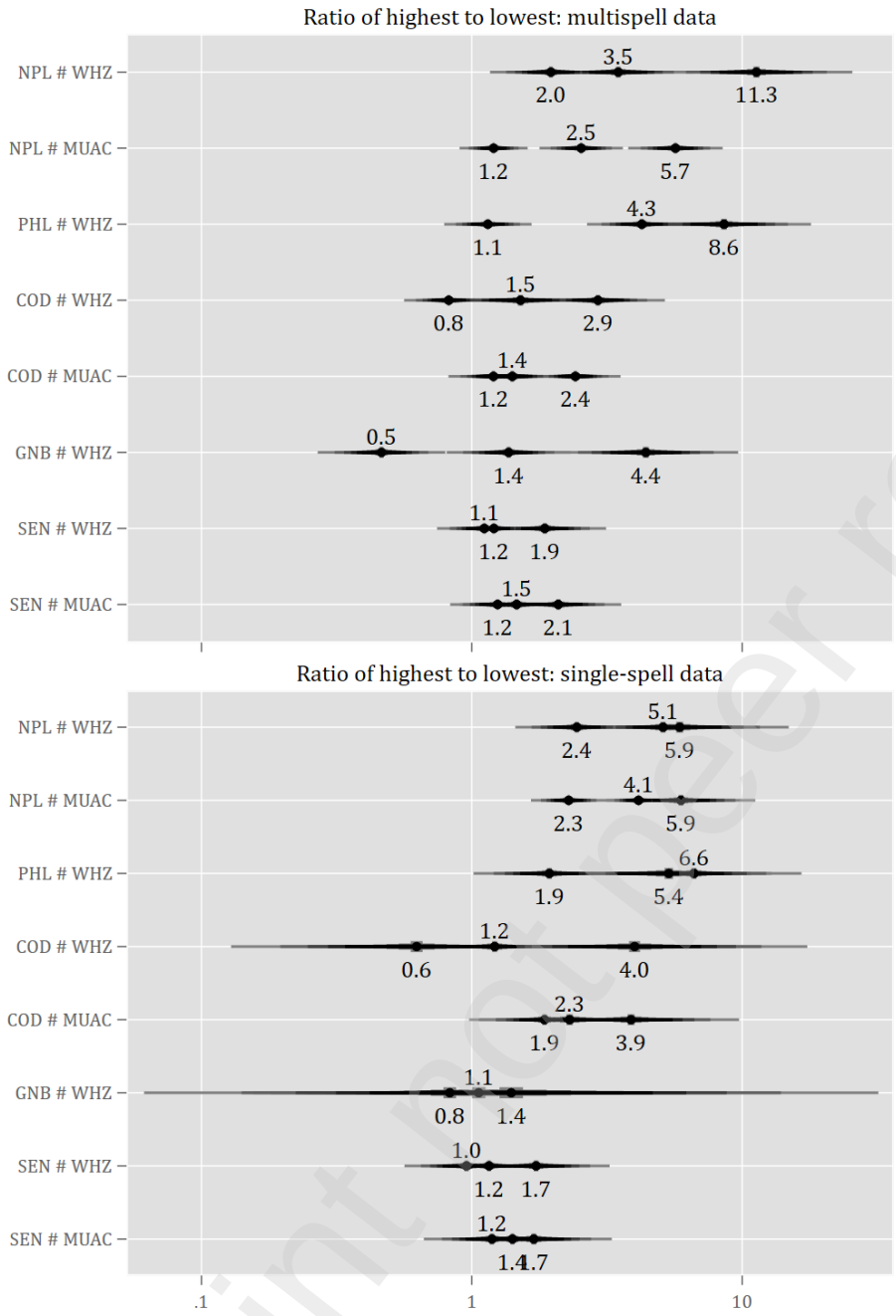


Figure 22. MIG model risk ratio estimates on multispell and single-spell data sets, including Olofin controls where available



5.4 MIG estimates

My preferred MIG specification incorporates the Olofin et al. controls, where defined, into the μ equation, and represents WHZ or MUAC quadratically in both the c and μ equations (sections 3.5 and 3.6 discussed these choices).

Table 4 shows the MIG parameter estimates when taking WHZ as the anthropometric indicator. Where Olofin controls are defined, regressions are reported both with and without them. But for clarity the coefficients on controls are not shown. The units of μ are changes in the health level per unit of time (the month), so the coefficients in the μ equation should also be taken per month.

Because WHZ enters four times—linearly and quadratically in two equations—the relevant coefficients are hard to interpret. At extreme low values of WHZ, such as -4, the quadratic terms dominate, and of these the one in the μ equation is usually larger, especially after multiplication by the average number of months in a spell. It is always positive, meaning that having extremely low (negative) WHZ is associated with greater risk. It is interesting, and unexpected, that the best fit ascribes

The consistent negative coefficients on age confirm that being older reduces mortality. Gender shows little consistent effect.

Since it is not immediately obvious how much a change in WHZ from, say –3 to –2, affects the risk of death, the last rows of the table return to risk ratios predicted by the fitted model, by way of equation (1). Two risk ratios are derived. First, to represent complete cure of SAM, risk ratios are once again computed for WHZ = –3.5 versus WHZ = –0.5. The predicted risk ratio ranges from 2.387 in Senegal to approximately 16 in the Philippines. These values are lower than the hazard ratios in Olofin et al. (2013, Table S2(C)), but different enough from each other to suggest that the effects of measurement aging have not been completely expunged. (Recall that the Philippines data features the most frequent follow-up and the Senegal data the least.)

The second comparison uses WHZ scores that are closer together, at –2.75 and –1.75, which may be more relevant for assessing the average impact of CMAM. As noted in the introduction, mean WHZ for children with WHZ < -2 in DHS data was found to be -2.77, and CMAM may be expected to increase WHZ for such children by one point on average. In each country, this narrowing naturally shrinks the mortality ratio; and it shrinks the standard errors even more since a larger denominator supports more stable ratios.

I perform the same analysis with MUAC in the place of WHZ. Since the WHO's recommended thresholds for diagnosing SAM and MAM are 11.5 and 12.5cm, higher than the corresponding WHZ thresholds by 14.5, for illustrative MUAC risk ratios, I add 14.5 to all the WHZ values used above. I compare MUAC = 11cm to MUAC = 14cm and MUAC = 11.75cm to MUAC = 12.75cm. See Table 5.

Table 4. Regression results for MIG model with initial health quadratic in WHZ

	DRC	Guinea-Bissau		Nepal		Philippines		Senegal
With controls?	No	No	Yes	No	Yes	No	Yes	No
c								
WHZ	0.250	0.279	0.277	1.110	0.952	0.478	0.514	0.007
	(0.116)	(0.196)	(0.195)	(0.429)	(0.365)	(0.309)	(0.382)	(0.171)
WHZ^2	0.038	0.004	0.003	0.137	0.110	0.069	0.066	−0.009
	(0.043)	(0.039)	(0.039)	(0.072)	(0.062)	(0.058)	(0.068)	(0.045)
Constant	1.796	1.784	1.783	3.354	3.157	1.844	1.919	2.412
	(0.189)	(0.407)	(0.407)	(0.597)	(0.495)	(0.395)	(0.512)	(0.328)
μ								
WHZ	0.082	0.099	0.092	0.126	0.146	0.205	0.228	−0.034
	(0.071)	(0.118)	(0.116)	(0.169)	(0.146)	(0.230)	(0.280)	(0.044)
WHZ^2	0.116	0.044	0.041	0.097	0.076	0.187	0.174	0.021
	(0.036)	(0.042)	(0.042)	(0.044)	(0.039)	(0.050)	(0.057)	(0.012)
Age (months)	−0.026	−0.016	−0.015	−0.012	−0.009	−0.040	−0.041	−0.012
	(0.004)	(0.007)	(0.007)	(0.005)	(0.005)	(0.014)	(0.017)	(0.002)
Female	−0.085	−0.134	−0.134	0.043	0.047	0.060	0.057	0.014
	(0.101)	(0.144)	(0.143)	(0.140)	(0.117)	(0.143)	(0.155)	(0.029)
$\ln \tau^2$								
Constant	0.563	0.173	0.157	0.697	0.228	0.848	0.866	−1.062
	(0.249)	(0.463)	(0.480)	(0.478)	(0.614)	(0.282)	(0.355)	(0.332)
Observations	17975	3274	3270	18309	18104	21888	19148	12610
One-year mortality ratio	6.120	2.971	2.942	9.161	8.738	17.402	15.645	2.387
WHZ = – 3.5 vs. – 0.5	(2.135)	(1.351)	(1.429)	(4.258)	(3.481)	(5.191)	(5.572)	(0.428)
One-year mortality ratio	2.003	1.518	1.509	2.383	2.359	3.045	2.967	1.376
WHZ = – 2.75 vs. – 1.75	(0.231)	(0.281)	(0.281)	(1.648)	(0.354)	(0.468)	(1.545)	(0.083)

Notes: Standard errors are in parenthesis, clustered by individual. Sample restricted to spells with initial age between 6 and 60 months. By the sign convention in the μ equation, negative coefficients cause higher values of corresponding variables to reduce the proximity of death, i.e., *increase* health. Control sets defined in Olofin et al. (2013), Table S1, included where available, in the μ equation; their coefficients are omitted for clarity. Mortality ratios are computed by equation (1), taking WHZ values as shown, t at 12 months, age at six months, and all other controls at their sample means. Standard errors of mortality ratios are computed via simulation: 10,000 draws are taken from the multivariate normal distribution implied by the parameter estimates and the log mortality ratio computed for each.

Table 5. Regression results for MIG model with initial health quadratic in MUAC

	DRC	Nepal		Senegal
With controls?	No	No	Yes	No
c				
MUAC (cm)	−0.262	0.208	0.214	0.536
	(0.976)	(0.417)	(0.415)	(0.641)
MUAC^2 (cm^2)	0.018	−0.004	−0.004	−0.018
	(0.040)	(0.019)	(0.019)	(0.025)
Constant	2.195	−0.421	−0.475	−1.577
	(5.884)	(2.207)	(2.193)	(4.207)
μ				
MUAC (cm)	−1.653	−1.667	−1.694	−0.416
	(0.589)	(0.322)	(0.325)	(0.199)
MUAC^2 (cm^2)	0.059	0.050	0.052	0.012
	(0.023)	(0.014)	(0.014)	(0.007)
Age (months)	−0.018	−0.002	−0.002	−0.010
	(0.004)	(0.002)	(0.002)	(0.002)
Female	−0.150	0.015	0.012	−0.018
	(0.095)	(0.059)	(0.059)	(0.029)
$\ln \tau^2$				
Constant	0.399	0.765	0.754	−1.087
	(0.243)	(0.147)	(0.146)	(0.389)
Observations	18295	104866	103848	12638
One-year mortality ratio	3.520	11.242	10.139	2.656
MUAC = 11cm vs. 14cm	(0.872)	(2.298)	(2.055)	(0.482)
One-year mortality ratio	1.618	2.338	2.270	1.408
MUAC = 11.75cm vs. 12.75cm	(0.140)	(0.220)	(0.226)	(0.138)

Notes: See notes to Table 4.

5.5 Meta-analysis

To blend results across data sets, and in a way that minimizes my discretion, I copy Olofin et al. in performing DerSimonian and Laird (1986) random-effects meta-analysis on the "with-controls-where-available" regressions. The meta-analysis is applied to the one-year risk ratios at the bottom of the regression tables, taken in logarithms.

Figure 23 and Figure 24 display the resulting forest plots for the two WHZ risk ratios. The meta-analytic bottom lines for the one-year mortality ratios are 5.61 for WHZ = −3.5 vs. −0.5 and 1.81 for WHZ = −2.75 vs. −1.75. The next two figures do the same for MUAC, and produce similar overall averages despite the changes in metric and sample: 4.56 for MUAC = 11cm vs. 14cm and 1.73 for MUAC = 11.75 vs. 12.75cm.

Figure 23. Forest plot for random-effects meta-analysis of WHZ = −3.5 vs. −0.5 mortality ratios in MIG model

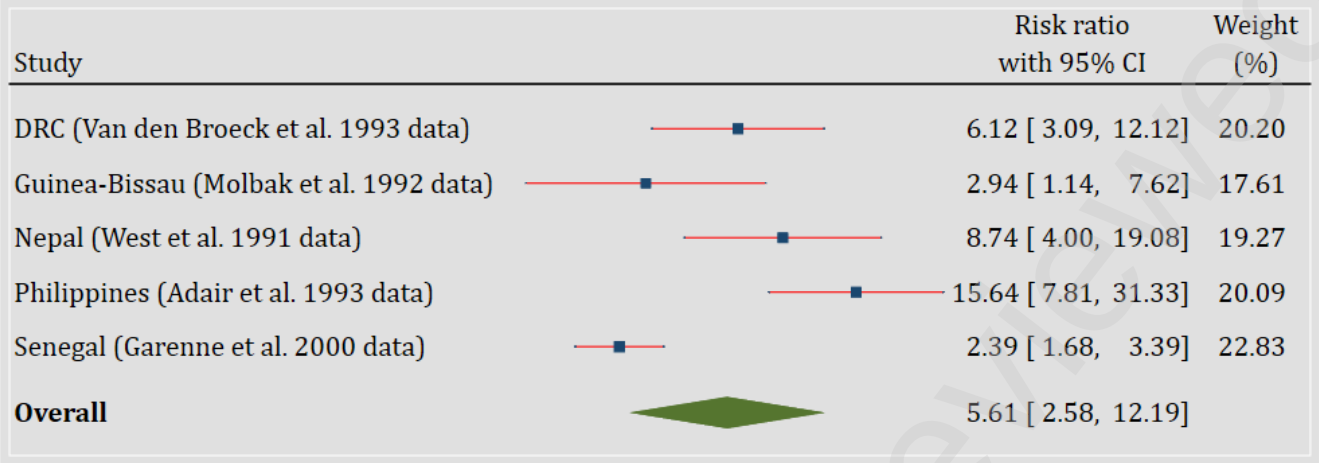


Figure 24. Forest plot for random-effects meta-analysis of WHZ = −2.75 vs. −1.75 mortality ratios in MIG model

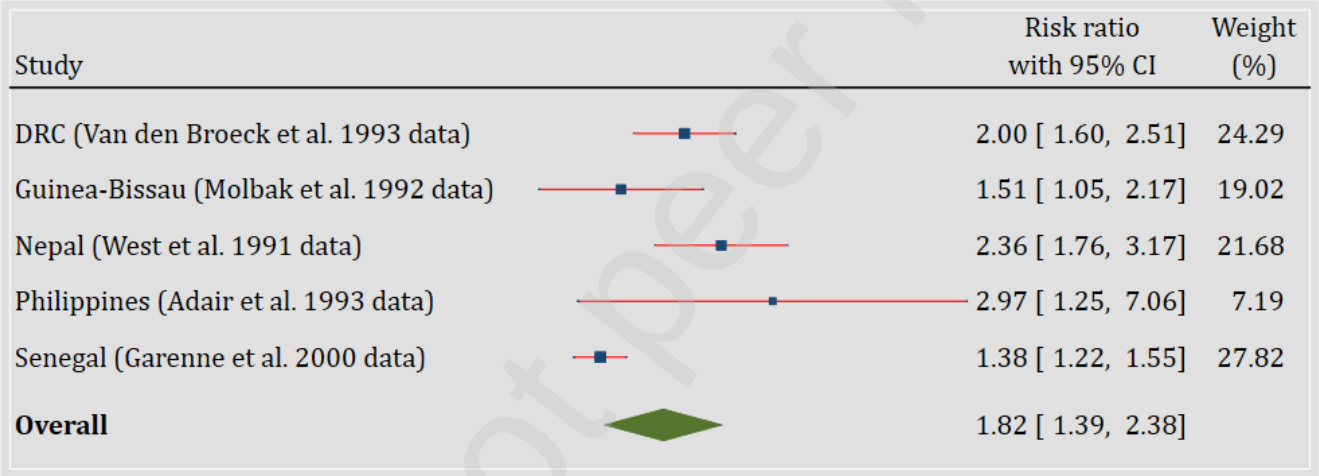


Figure 25. Forest plot for random-effects meta-analysis of MUAC = 11cm vs. 14cm mortality ratios in MIG model

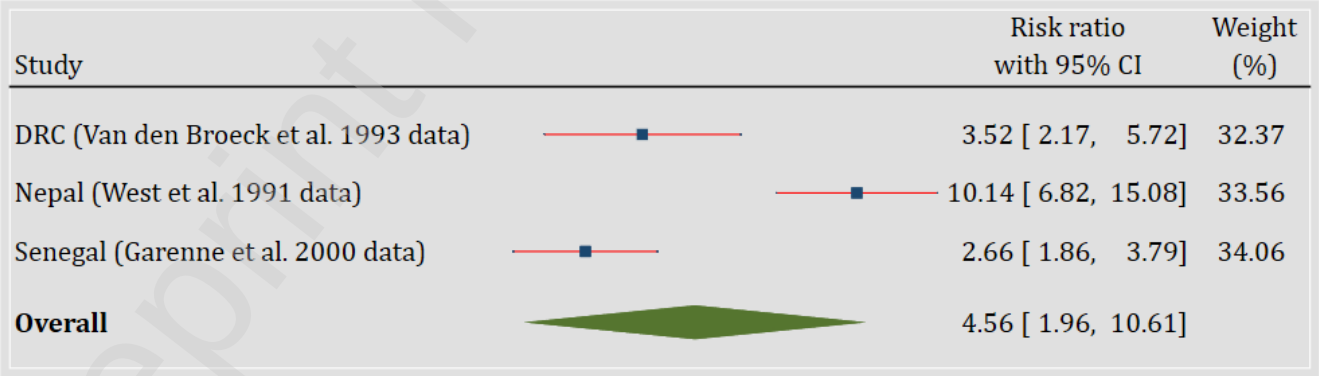
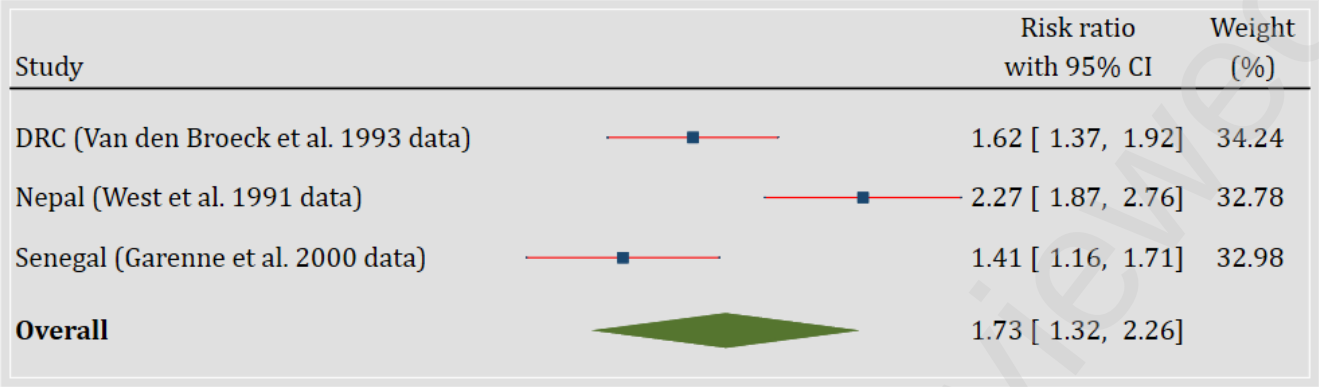


Figure 26. Forest plot for random-effects meta-analysis of MUAC = 11.75cm vs. 12.75cm mortality ratios in MIG model



6 Probit results

To check the MIG results I return to the simpler approach that emerged in the analytical narrative: a probit model fit to single-spell data sets. As in the MIG modeling, I enter anthropometric variables quadratically, control for age and gender, and include the Olofin controls where available. Since timing information for death is not needed, the Indonesia data are now incorporated.

Parameter estimates and model-based one-year risk ratios appear in Table 6 and Table 7. The forest plots for WHZ are in Figure 27 and Figure 28 and those for MUAC in Figure 29 and Figure 30. All follow the formats of corresponding MIG displays. The meta-analytic bottom lines change remarkably little, coming modestly lower. The mortality ratio for the narrower comparison is 1.58 for WHZ and 1.56 for MUAC. Given all the differences in sample, data structure, and method, this concordance is reassuring.

Table 6. Regression results for probit model with initial health quadratic in WHZ

	DRC	Guinea-Bissau		Indonesia		Nepal		Philippines		Senegal
With controls?	No	No	Yes	No	Yes	No	Yes	No	Yes	No
WHZ	−0.097	−0.054	−0.033	−0.235	−0.253	−0.218	−0.200	−0.144	−0.141	−0.060
	(0.050)	(0.084)	(0.086)	(0.061)	(0.064)	(0.119)	(0.121)	(0.065)	(0.073)	(0.038)
WHZ^2	0.008	0.067	0.096	−0.033	−0.040	0.028	0.025	0.051	0.048	0.034
	(0.023)	(0.035)	(0.040)	(0.027)	(0.029)	(0.030)	(0.031)	(0.025)	(0.028)	(0.015)
Age (months)	−0.012	−0.157	−0.187	0.015	0.022	−0.042	−0.042	−0.012	−0.020	0.021
	(0.089)	(0.178)	(0.184)	(0.085)	(0.088)	(0.112)	(0.116)	(0.106)	(0.119)	(0.064)
Female	−0.022	−0.011	−0.011	−0.010	−0.008	−0.009	−0.009	0.005	−0.003	−0.021
	(0.003)	(0.011)	(0.011)	(0.003)	(0.004)	(0.005)	(0.005)	(0.054)	(0.061)	(0.002)
Observations	4464	529	477	2642	2517	3349	2986	2566	2223	4933
One-year mortality ratio	2.097	4.454	6.159	1.715	1.642	8.400	7.184	7.664	7.732	2.546
WHZ = – 3.5 vs. – 0.5	(0.831)	(2.003)	(2.920)	(0.673)	(0.727)	(2.725)	(2.443)	(2.595)	(3.059)	(0.456)
One-year mortality ratio	1.287	1.755	2.019	1.155	1.131	2.075	1.968	2.078	2.082	1.406
WHZ = – 2.75 vs. – 1.75	(0.196)	(0.322)	(0.405)	(0.170)	(0.187)	(0.254)	(0.252)	(0.289)	(0.336)	(0.100)

Notes: Standard errors are in parenthesis, clustered by individual. Samples consist of one one-year spell per child, starting from first measurement after age 6 months. Control sets defined in Olofin et al. (2013), Table S1, included where available; their coefficients are omitted for clarity. In computing mortality ratios, WHZ values are taken as shown, age is taken at six months, and all controls are taken at their sample means. Standard errors of mortality ratios are computed via simulation: 10,000 draws are taken from the multivariate normal distribution implied by the parameter estimates and the log mortality ratio computed for each.

Table 7. Regression results for probit model with initial health quadratic in MUAC

	DRC	Nepal		Senegal
With controls?	No	No	Yes	No
MUAC (cm)	−0.408	−1.235	−1.270	−0.675
	(0.478)	(0.175)	(0.177)	(0.298)
$MUAC^2$ (cm^2)	0.010	0.038	0.040	0.019
	(0.019)	(0.007)	(0.007)	(0.011)
Age (months)	−0.054	−0.017	−0.021	−0.038
	(0.089)	(0.046)	(0.047)	(0.064)
Female	−0.017	−0.004	−0.004	−0.019
	(0.004)	(0.002)	(0.002)	(0.003)
Observations	4573	18875	18197	4940
One-year mortality ratio	2.585	6.644	6.182	2.585
MUAC = 11cm vs. 14cm	(0.688)	(0.861)	(0.829)	(0.393)
One-year mortality ratio	1.385	1.964	1.926	1.391
MUAC = 11.75cm vs. 12.75cm	(0.119)	(0.081)	(0.082)	(0.080)

Figure 27. Forest plot for random-effects meta-analysis of WHZ = −3.5 vs. WHZ = −0.5 mortality ratios in probit model

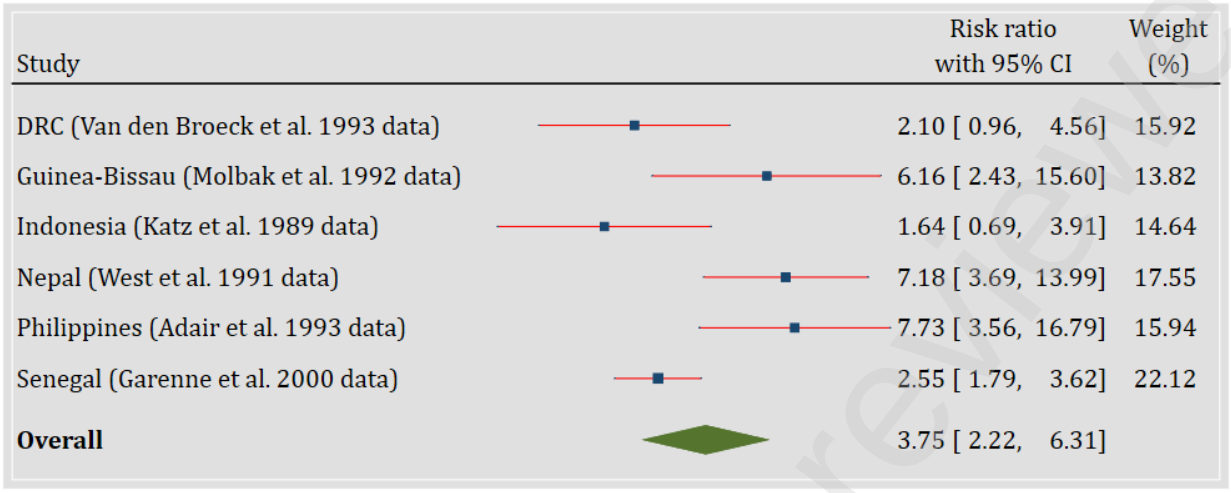


Figure 28. Forest plot for random-effects meta-analysis of WHZ = −2.75 vs. WHZ = −1.75 mortality ratios in probit model

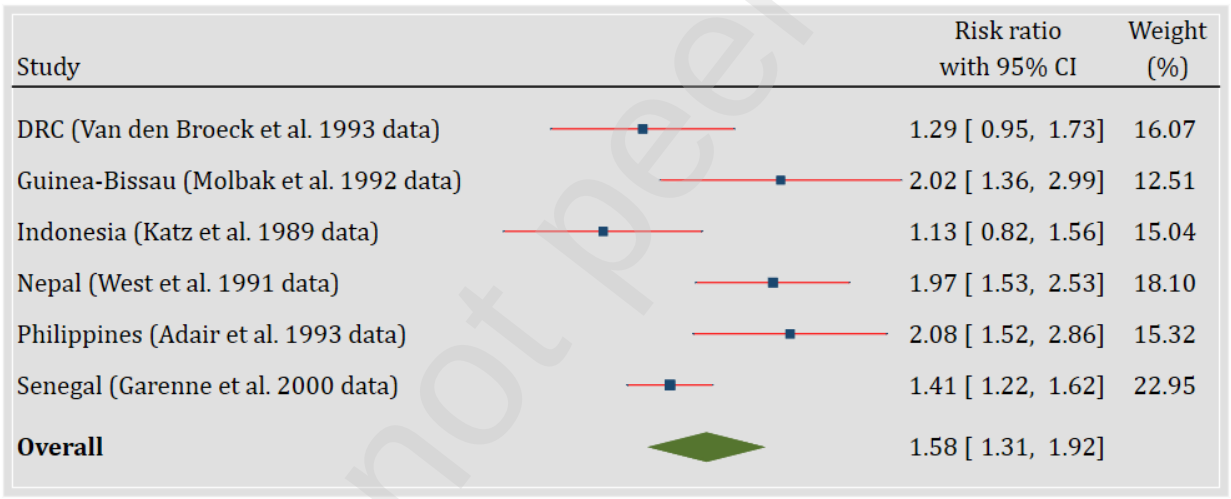


Figure 29. Forest plot for random-effects meta-analysis of MUAC = 11 vs. MUAC = 14 mortality ratios in probit model

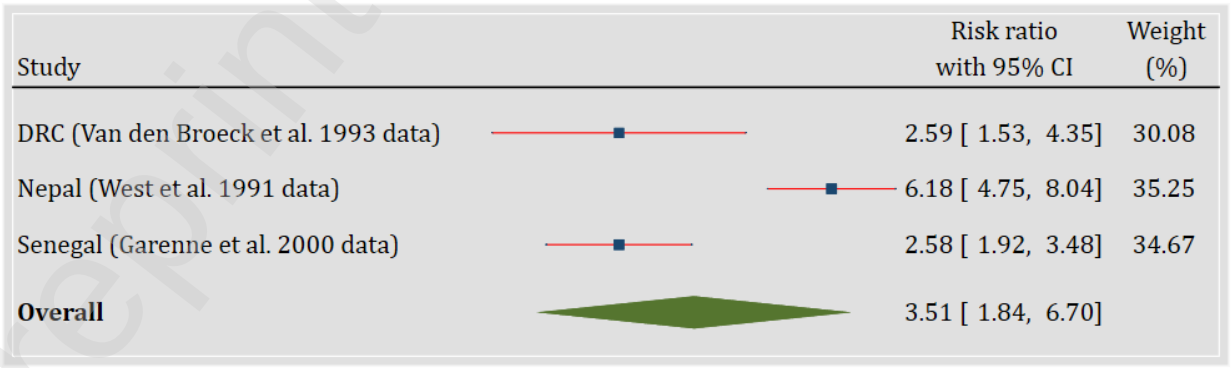
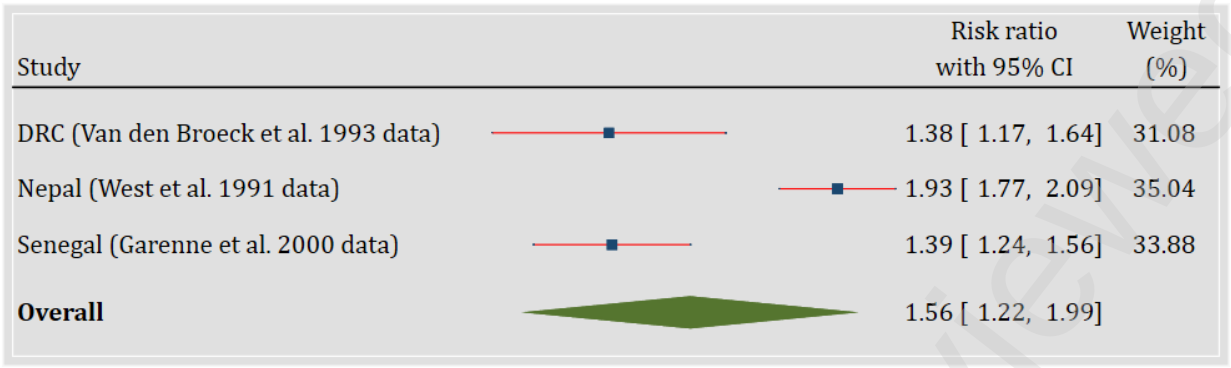


Figure 30. Forest plot for random-effects meta-analysis of MUAC = 11.75 vs. MUAC = 12.75 mortality ratios in probit model



7 From statistical modeling to predictive modeling

The purpose of the estimation in the previous two sections is to *understand the process that generated the data that we have*. This is distinct from *predicting the impact of an intervention*. As far as we know, the intervention is not present in our data.

The question of how best to bridge from the first to second is hard. If we learn that in some target setting today, the average child diagnosed with SAM exits CMAM with WHZ $= -1.5$, it does not follow that the child will face, and have just faced, the same mortality risk as children with WHZ $= -1.5$ in the data analyzed above.

GiveWell Senior Researcher Stephan Guyenet and I collaborated on a preliminary methodology to incorporate the observational risk ratio estimates above into an estimate of the impact of CMAM.

The methodology incorporates MIG model results in two places:

1. To estimate the counterfactual mortality ratio of children with SAM or MAM who would be admitted to the treatment program, relative to non-malnourished children. In other words, in the absence of treatment, how elevated is the risk of death among malnourished children over a one-year period, relative to non-malnourished children?
2. To estimate the impact of the intervention. In other words, among malnourished children, how much does the intervention reduce the risk of dying over a one-year period, relative to the counterfactual of not receiving treatment? We separately estimate the impact of government-only treatment and NGO-supported treatment.

The CEA uses these ratios to estimate the all-cause mortality impact of transitioning from no treatment to NGO-supported treatment, and transitioning from government treatment to NGO-supported treatment.

Conceptually, the method for generating the ratios begins with measured WHZ of children in MAM, SAM and non-malnourished categories in the countries of interest. It uses these WHZ inputs to generate country-specific annual mortality ratios for untreated SAM and MAM relative to non-malnourished children (item 1 above). To estimate a treatment effect (2), the method begins by estimating WHZ gain due to CMAM. It then uses the MIG model to generate annual mortality ratios based on baseline vs. post-treatment WHZ for MAM and SAM.

The method elaborates the process presented in the previous two sections. Instead of computing a single risk ratio such as for WHZ = -3.5 vs. -0.5, it computes distinct ratios for thousands of children whose data appear in Demographic and Health Surveys. For example, for an eight-month-

old girl appearing in the Chad data with WHZ $= -3.5$, which qualifies for SAM diagnosis, it is assumed that NGO-administered CMAM will increase WHZ by 1.54, bringing the child to WHZ = -1.96. The risk ratio is then estimated for WHZ = -3.5 vs. -1.96. The thousands of (log) risk ratios are then averaged. This elaboration applies after the MIG model fitting, so that all risk ratios come from the MIG estimates in section 5, but before meta-analysis.

To be precise, the process runs as follows:

- Seven prospective intervention countries are identified: Burkina Faso, Chad, the DRC, Ethiopia, Mali, Niger, and Nigeria.
- The most recent DHS microdata from those countries are obtained. Age, gender, and WHZ observations are retained for all children aged 6–60 months whose WHZ classifies them as SAM or MAM (WHZ < -2).[26]
- Three comparison groups are constructed, one to compute the malnourished-nourished risk ratios in item 1 above and two for the untreated-treated ratios in item 2 (for government- and NGO-run treatment). For the first, the comparator is fixed at the country's average WHZ for those with WHZ > 2, according to the DHS data. For the rest, the "treatment groups" are constructed by adding category-specific increments to each malnourished child's WHZ. The increments are differentiated, by country, by SAM vs. MAM status, and by government vs. NGO administrator type. They are derived from literature using methods described below. The estimates are very similar across countries and agency types. The sample means of these WHZ values are collated in the "WHZ low" and "WHZ high" columns of Table 9 and Table 10.
- We do not currently have direct estimates of WHZ at admission and discharge from the programs we are evaluating. To estimate mean WHZ gain resulting from CMAM treatment, we performed a literature review of the impact of CMAM programs on WHZ and took a weighted average. The results suggest that CMAM increases WHZ by 0.8 in children with MAM and by 1.3 in children with SAM. We then further adjusted these values to account for apparently greater treatment success rates in NGO-supported vs. government-only CMAM programs. After this adjustment, NGO-supported CMAM programs are estimated to increase mean WHZ by about 0.8 in children with MAM and about 1.5 in children with SAM. (See Table 8.)
- For each DHS observation, each of the three comparisons, and each of the five historical data sets used in the MIG WHZ estimation, the one-year log mortality ratios are computed, just as in Table 4.
- Using DHS sampling weights, the resulting individual-level log mortality ratios are averaged across DHS samples, for each historical data set, producing point estimates analogous to the one-year risk ratios at the bottom of the regression tables above—for each historical data set, intervention country, and agency type.
- To generate standard errors for each ratio, this process is repeated 10,000 times, each iteration taking a random draw from the multivariate normal parameter distribution implied by a country's estimation result as summarized in Table 4.[27]
- To combine results across historical data sets, producing one estimate for each target country and agency type, random-effects meta-regression is performed, as above.

[26] To my knowledge, DHS surveys do not capture MUAC, and the same goes for UNICEF's MICS. SMART surveys do emphasize MUAC but the SMART program does not appear from its website to share microdata as readily.

[27] The variance of the normal distribution is the full parameter covariance matrix.

The final risk ratios are in Table 9 and Table 10. The major differences are across the SAM/MAM divide. Inverting the ratios in Table 10, the analysis roughly suggests that CMAM reduces one-year mortality by 40% among children with MAM and by 70% among children with SAM.

This predictive modeling approach assumes that the effect of malnutrition treatment on mortality is captured by its impact on WHZ. We believe this assumption is highly uncertain, and likely incorrect. There are at least three specific reasons to believe the mortality impact of CMAM treatment programs may diverge from our current estimates:

- CMAM treatment programs include micronutrient supplementation, which could impact mortality risk independently of WHZ. For example, vitamin A is a component of RUTF/RUSF supplements, and vitamin A supplementation reduces mortality by reducing susceptibility to certain infectious diseases. This would cause our current model to underestimate the treatment effect.
- CMAM treatment programs include antibiotic treatment, which could impact mortality risk independently of WHZ. This would cause our current model to underestimate the treatment effect.
- Elevated mortality risk associated with malnutrition in the data sets underlying the MIG model may not be entirely causally attributable to malnutrition *per se*. It may be partially a result of confounding by socioeconomic conditions or other factors. This would cause our current model to overestimate the treatment effect.

An additional source of uncertainty is the WHZ values at admission and discharge, which are not based on data from the programs themselves.

We currently use this method as a rough approximation in our cost-effectiveness model as we refine our estimation method. We will likely use the method described and refine it by adding adjustments for important variables, such as those described above, rather than starting from scratch. However, we remain open to alternative methods.

Table 8. WHZ increments used to estimate mortality impact of community-based management of malnutrition (CMAM)

	Burkina Faso	Chad	DRC	Ethiopia	Mali	Niger	Nigeria
Mean WHZ gain from government treatment, MAM	0.77	0.77	0.77	0.77	0.77	0.77	0.77
Mean WHZ gain from government treatment, SAM	1.27	1.27	1.27	1.27	1.27	1.27	1.27
Mean WHZ gain from NGO treatment, MAM	0.78	0.78	0.78	0.80	0.81	0.78	0.78
Mean WHZ gain from NGO treatment, SAM	1.50	1.62	1.50	1.50	1.50	1.44	1.54

Table 9. Meta-analytical one-year mortality ratios for malnourished vs. well-nourished children

	Burkina Faso			Chad			DRC			Ethiopia			Mali			Niger			Nigeria		
	WHZ low	WHZ high	Risk ratio	WHZ low	WHZ high	Risk ratio	WHZ low	WHZ high	Risk ratio	WHZ low	WHZ high	Risk ratio	WHZ low	WHZ high	Risk ratio	WHZ low	WHZ high	Risk ratio	WHZ low	WHZ high	Risk ratio
MAM	−2.41 (0.28)	−0.30 0.00	2.51 (0.51)	−2.42 (0.28)	−0.25 0.00	2.58 (0.54)	−2.44 (0.29)	−0.07 0.00	2.61 (0.54)	−2.31 (0.25)	−0.20 0.00	2.44 (0.47)	−2.40 (0.28)	−0.35 0.00	2.48 (0.50)	−2.42 (0.28)	−0.52 0.00	2.54 (0.53)	−2.39 (0.29)	−0.16 0.00	2.44 (0.48)
SAM	−3.69 (0.47)	−0.30 0.00	7.12 (3.01)	−3.62 (0.47)	−0.25 0.00	7.01 (2.92)	−3.72 (0.50)	−0.07 0.00	7.60 (3.25)	−3.53 (0.40)	−0.20 0.00	7.07 (2.93)	−3.70 (0.55)	−0.35 0.00	7.18 (3.07)	−3.71 (0.55)	−0.52 0.00	6.90 (2.89)	−3.57 (0.45)	−0.16 0.00	6.39 (2.55)

Note: WHZ values are means of DHS samples. Standard deviations (for WHZ low) and standard errors (for risk ratios) in parentheses. "WHZ high" values have zero standard deviation because the exact value shown is used for every subject in the samples.

Table 10. Meta-analytical one-year mortality ratios for proxy treatment vs proxy control groups

	Burkina Faso			Chad			DRC			Ethiopia			Mali			Niger			Nigeria		
	WHZ low	WHZ high	Risk ratio	WHZ low	WHZ high	Risk ratio	WHZ low	WHZ high	Risk ratio	WHZ low	WHZ high	Risk ratio	WHZ low	WHZ high	Risk ratio	WHZ low	WHZ high	Risk ratio	WHZ low	WHZ high	Risk ratio
MAM, gov.	−2.41 (0.28)	−1.64 (0.28)	1.68 (0.20)	−2.42 (0.28)	−1.65 (0.28)	1.69 (0.20)	−2.44 (0.29)	−1.67 (0.29)	1.70 (0.21)	−2.31 (0.25)	−1.55 (0.25)	1.68 (0.20)	−2.40 (0.28)	−1.63 (0.28)	1.67 (0.20)	−2.42 (0.28)	−1.65 (0.28)	1.70 (0.20)	−2.39 (0.29)	−1.62 (0.29)	1.67 (0.19)
SAM, gov.	−3.69 (0.47)	−2.42 (0.47)	2.82 (0.64)	−3.62 (0.47)	−2.35 (0.47)	2.89 (0.67)	−3.72 (0.50)	−2.46 (0.50)	2.91 (0.67)	−3.53 (0.40)	−2.27 (0.40)	2.98 (0.71)	−3.70 (0.55)	−2.43 (0.55)	2.84 (0.65)	−3.71 (0.55)	−2.44 (0.55)	2.85 (0.64)	−3.57 (0.45)	−2.30 (0.45)	2.78 (0.64)
MAM, NGO	−2.41 (0.28)	−1.63 (0.28)	1.69 (0.20)	−2.42 (0.28)	−1.63 (0.28)	1.71 (0.21)	−2.44 (0.29)	−1.66 (0.29)	1.72 (0.21)	−2.31 (0.25)	−1.52 (0.25)	1.71 (0.21)	−2.40 (0.28)	−1.59 (0.28)	1.71 (0.21)	−2.42 (0.28)	−1.63 (0.28)	1.71 (0.21)	−2.39 (0.29)	−1.61 (0.29)	1.68 (0.20)
SAM, NGO	−3.69 (0.47)	−2.19 (0.47)	3.34 (0.89)	−3.62 (0.47)	−2.00 (0.47)	3.72 (1.07)	−3.72 (0.50)	−2.22 (0.50)	3.47 (0.94)	−3.53 (0.40)	−2.03 (0.40)	3.55 (0.98)	−3.70 (0.55)	−2.20 (0.55)	3.37 (0.90)	−3.71 (0.55)	−2.27 (0.55)	3.23 (0.82)	−3.57 (0.45)	−2.03 (0.45)	3.36 (0.92)

Note: WHZ values are means of DHS samples. Standard deviations (for WHZ low and WHZ high) and standard errors (for risk ratios) in parentheses.

Acknowledgments

Deep thanks to Linda Adair, Michel Garenne, Joanne Katz, Kåre Mølbak, Keith West, and Lee Shu–Fune Wu for sharing data, and to Megan Higgs and GiveWell colleagues for transformational feedback. Many of the graphs are made with coefplot (Jann 2014) and plottig (Bischof 2017).

References

Aalen, Odd O. 1994. "Effects of Frailty in Survival Analysis." *Statistical Methods in Medical Research* 3 (3): 227–43. doi.org/10.1177/096228029400300303.

Aalen, Odd O., Ørnulf Borgan, and Håkon K. Gjessing. 2008. *Survival and Event History Analysis*. doi.org/10.1007/978–0–387–68560–1.

Adair, L., B. M. Popkin, J. VanDerslice, J. Akin, D. Guilkey, R. Black, J. Briscoe, and W. Flieger. 1993. "Growth Dynamics during the First Two Years of Life: A Prospective Study in the Philippines." *European Journal of Clinical Nutrition* 47 (1): 42–51. ncbi.nlm.nih.gov/pubmed/8422872.

Bischof, Daniel. 2017. "New Graphic Schemes for Stata: Plotplain and Plottig." *The Stata Journal* 17 (3): 748–59. doi.org/10.1177/1536867X1701700313.

Black, Robert E., Lindsay H. Allen, Zulfiqar A. Bhutta, Laura E. Caulfield, Mercedes de Onis, Majid Ezzati, Colin Mathers, and Juan Rivera. 2008. "Maternal and Child Undernutrition: Global and Regional Exposures and Health Consequences." *The Lancet* 371 (9608): 243–60. doi.org/10.1016/S0140–6736(07)61690–0.

Daures, Maguy, Kevin Phelan, Mariama Issoufou, Séni Kouanda, Ousmane Sawadogo, Kader Issaley, Cecile Cazes, et al. 2020. "New Approach to Simplifying and Optimising Acute Malnutrition Treatment in Children Aged 6-59 Months: The OptiMA Single-Arm Proof-of-Concept Trial in Burkina Faso." *The British Journal of Nutrition* 123 (7): 756–67. doi.org/10.1017/S0007114519003258.

DerSimonian, R., and N. Laird. 1986. "Meta-Analysis in Clinical Trials." *Controlled Clinical Trials* 7 (3): 177–88. doi.org/10.1016/0197–2456(86)90046–2.

Fishman, Steven M., et al. "Childhood and Maternal Underweight." *Comparative Quantification of Health Risks: Global and Regional Burden of Disease Attributable to Selected Major Risk Factors*, vol. 1, 2004, pp. 39–161, google.com/books/edition/Comparative_Quantification_of_Health_Ris/ACV1jEGx4AgC?gbpv=1&pg=PA39.

Garenne, Michel, Bernard Maire, Olivier Fontaine, Khady Dieng, and André Briend. 2000. *Risques de Décès Associés: À Différents États Nutritionnels chez l'Enfant d'Âge Préscolaire.* Centre Population et Développement. documentation.ird.fr/hor/fdi:010023180.

Grellety, Emmanuel, and Michael H. Golden. 2016. "The Effect of Random Error on Diagnostic Accuracy Illustrated with the Anthropometric Diagnosis of Malnutrition." *PloS One* 11 (12): e0168585. doi.org/10.1371/journal.pone.0168585.

Grellety, Emmanuel, and Michael Golden. 2018. "Severely Malnourished Children with a Low Weight-for-Height Have a Similar Mortality to Those with a Low Mid-Upper-Arm-Circumference: II. Systematic Literature Review and Meta-Analysis," *Nutrition Journal* 17, 80. doi.org/10.1186/s12937–018–0383–5.

Hurn, Aubrey, Joseph Jeisman, and Kenneth Lindsay. 2007. "Seeing the Wood for the Trees: A Critical Evaluation of Methods to Estimate the Parameters of Stochastic Differential Equations." *Journal of Financial Econometrics* 5 (3): 390–455. doi.org/10.1093/jjfinec/nbm009.

International Rescue Committee. Undated. *Simplified Protocol for the Treatment of Uncomplicated Severe Acute Malnutrition Integrated into iCCM*. jogh.org/documents/issue202001/jogh–10–010421–s001.pdf

Jann, Ben. 2014. "Plotting Regression Coefficients and Other Estimates." *The Stata Journal* 14 (4): 708–37. doi.org/10.1177/1536867X1401400402.

Katz, J., K. P. West Jr, I. Tarwotjo, and A. Sommer. 1989. "The Importance of Age in Evaluating Anthropometric Indices for Predicting Mortality." *American Journal of Epidemiology* 130 (6): 1219–26. doi.org/10.1093/oxfordjournals.aje.a115450.

Mølbak, Kåre, Peter Aaby, Liselotte Ingholt, Niels Højlyng, Adam Gottschau, Henning Andersen, Lene Brink, et al. 1992. "Persistent and Acute Diarrhoea as the Leading Causes of Child Mortality in Urban Guinea Bissau." *Transactions of the Royal Society of Tropical Medicine and Hygiene*. doi.org/10.1016/0035–9203(92)90580–6.

Olofin, Ibironke, Christine M. McDonald, Majid Ezzati, Seth Flaxman, Robert E. Black, Wafaie W. Fawzi, Laura E. Caulfield, Goodarz Danaei, and for the Nutrition Impact Model Study (anthropometry cohort pooling). 2013. "Associations of Suboptimal Growth with All-Cause and Cause-Specific Mortality in Children under Five Years: A Pooled Analysis of Ten Prospective Studies." *PLoS ONE*. doi.org/10.1371/journal.pone.0064636

Pelletier, D. L., E. A. Frongillo Jr, D. G. Schroeder, and J. P. Habicht. 1994. "A Methodology for Estimating the Contribution of Malnutrition to Child Mortality in Developing Countries." *The Journal of Nutrition* 124 (10 Suppl): 2106S–2122S. doi.org/10.1093/jn/124.suppl_10.2106S.

Schwinger, Catherine, Michael H. Golden, Emmanuel Grellety, Dominique Roberfroid, and Benjamin Guesdon. 2019. "Severe Acute Malnutrition and Mortality in Children in the Community: Comparison of Indicators in a Multi-Country Pooled Analysis." *PloS One* 14 (8): e0219745. doi.org/10.1371/journal.pone.0219745.

Turnbull, B. W. 1976. "The empirical distribution function with arbitrarily grouped, censored and truncated data." *Journal of the Royal Statistical Society, Series B* 38: 290–95. doi.org/10.1111/j.2517–6161.1976.tb01597.x.

Van den Broeck, J., R. Eeckels, and J. Vuylsteke. 1993. "Influence of Nutritional Status on Child Mortality in Rural Zaire." *The Lancet* 341 (8859): 1491–95. doi.org/10.1016/0140–6736(93)90632–Q.

West, K. P., Jr, J. Katz, S. C. LeClerq, and E. K. Pradhan. 1991. "Efficacy of Vitamin A in Reducing Preschool Child Mortality in Nepal." *The Lancet*. doi.org/10.1016/0140-6736(91)90070-6.

World Health Organization (WHO). 2006. *WHO Child Growth Standards: Length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age*. Geneva. who.int/iris/bitstream/handle/10665/43413/924154693X_eng.pdf.

WHO. 2013. *Guideline: Updates on the Management of Severe Acute Malnutrition in Children.* https://apps.who.int/iris/bitstream/handle/10665/95584/9789241506328_eng.pdf.

WHO and UNICEF. 2009. *WHO child growth standards and the identification of severe acute malnutrition in infants and children: A joint statement by the World Health Organization and the United Nations Children's fund.* ncbi.nlm.nih.gov/books/NBK200776.

Appendix. Sensitivity test

The following displays show how the results in Figure 23–Figure 30, Table 9, and Table 10 change when the timeframe for impact is reduced from 12 to 3 months.

Figure 31. Forest plot for random-effects meta-analysis of WHZ = −3.5 vs. −0.5 mortality ratios in MIG model, 3-month follow-up

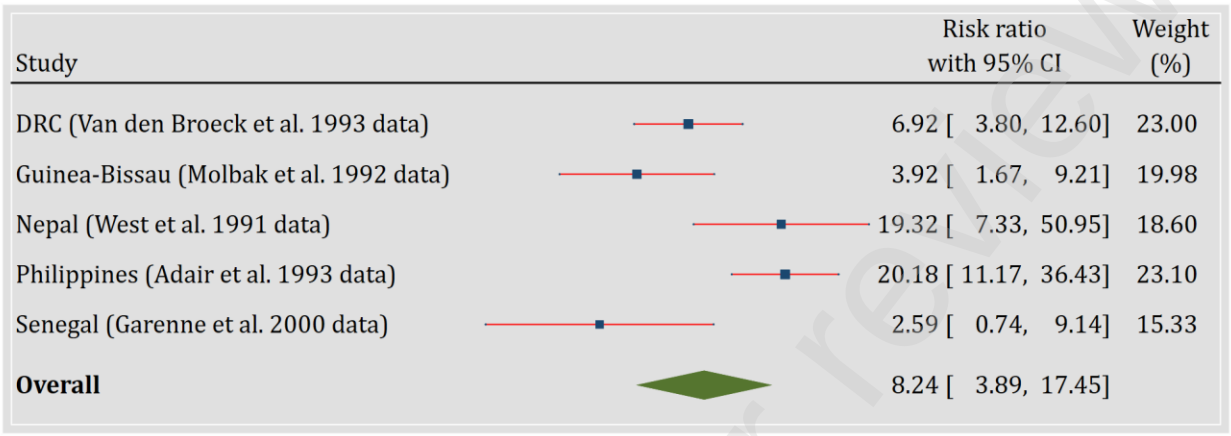


Figure 32. Forest plot for random-effects meta-analysis of WHZ = −2.75 vs. −1.75 mortality ratios in MIG model, 3-month follow-up

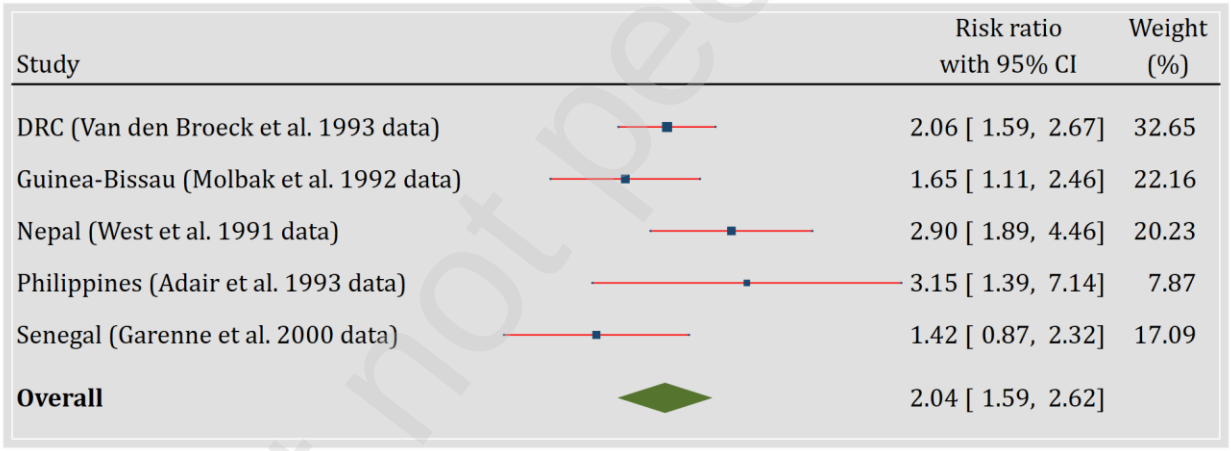


Figure 33. Forest plot for random-effects meta-analysis of MUAC = 11cm vs. 14cm mortality ratios in MIG model, 3-month follow-up

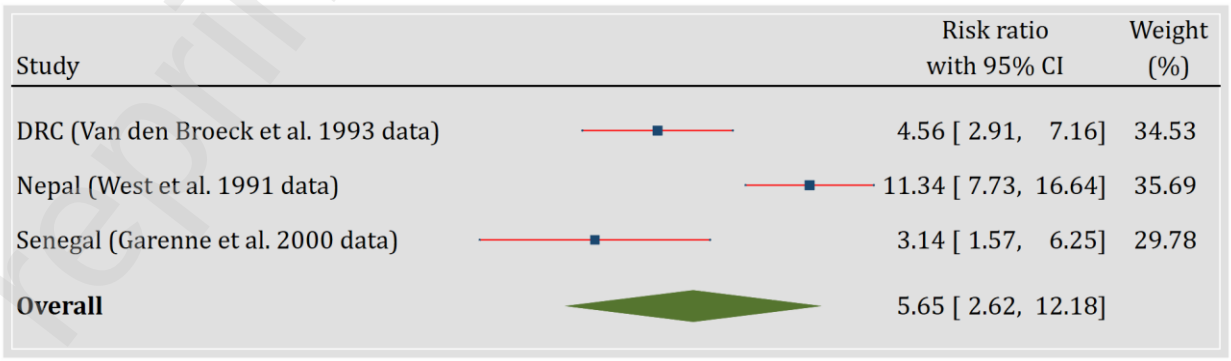


Figure 34. Forest plot for random-effects meta-analysis of MUAC = 11.75cm vs. 12.75cm mortality ratios in MIG model, 3-month follow-up

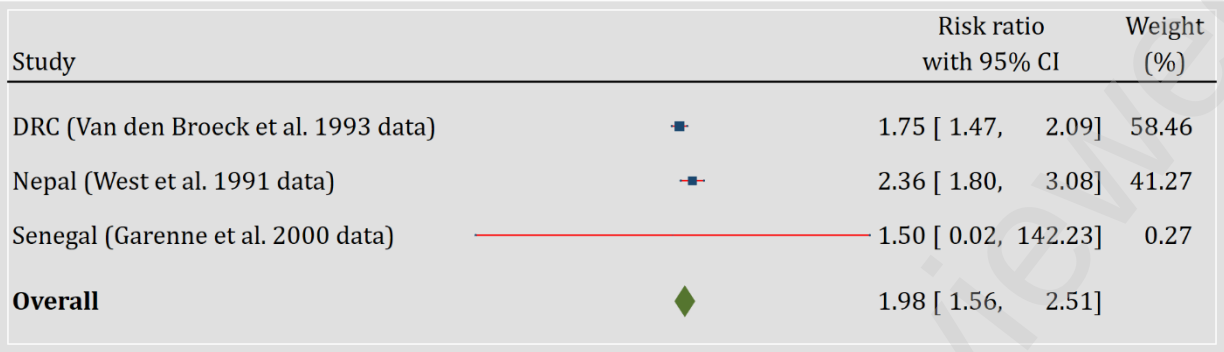


Figure 35. Forest plot for random-effects meta-analysis of WHZ = −3.5 vs. WHZ = −0.5 mortality ratios in probit model, 3-month follow-up

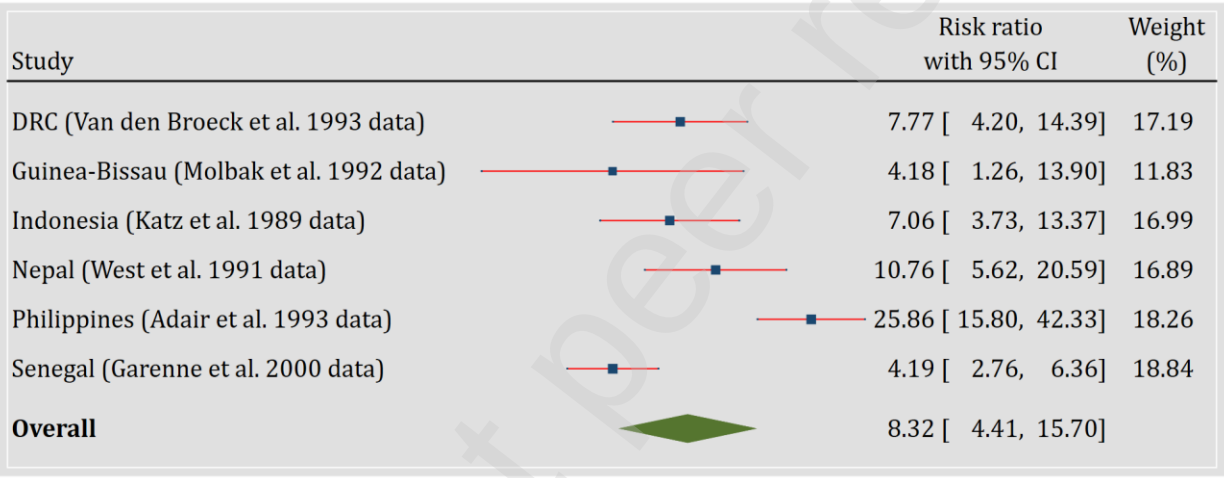


Figure 36. Forest plot for random-effects meta-analysis of WHZ = −2.75 vs. WHZ = −1.75 mortality ratios in probit model, 3-month follow-up

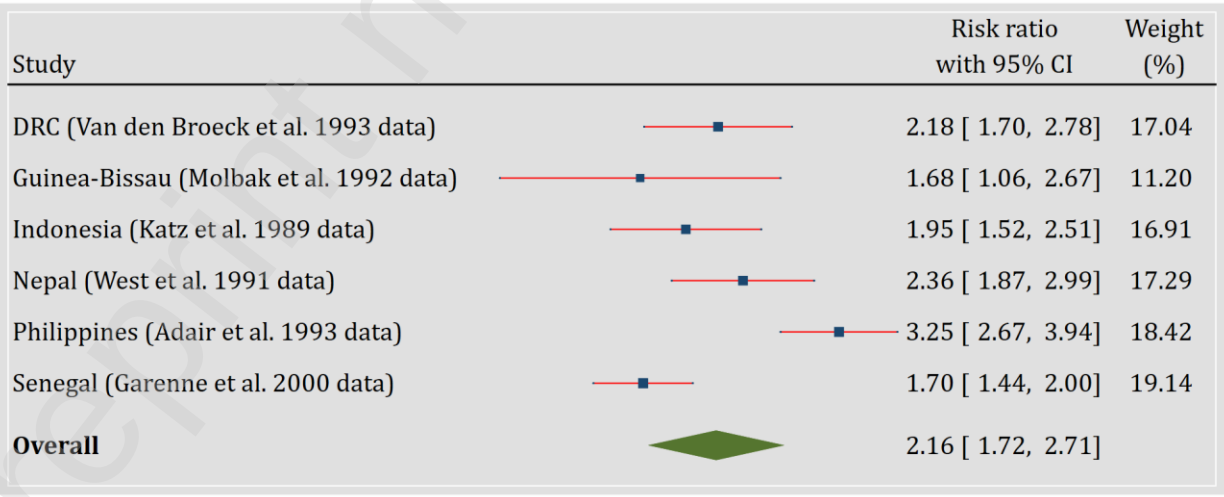


Figure 37. Forest plot for random-effects meta-analysis of MUAC = 11 vs. MUAC = 14 mortality ratios in probit model, 3-month follow-up

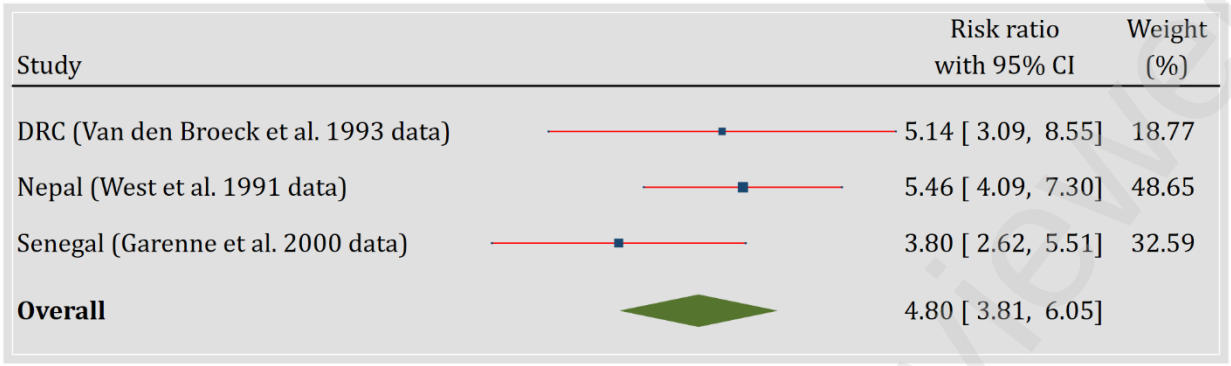


Figure 38. Forest plot for random-effects meta-analysis of MUAC = 11.75 vs. MUAC = 12.75 mortality ratios in probit model, 3-month follow-up

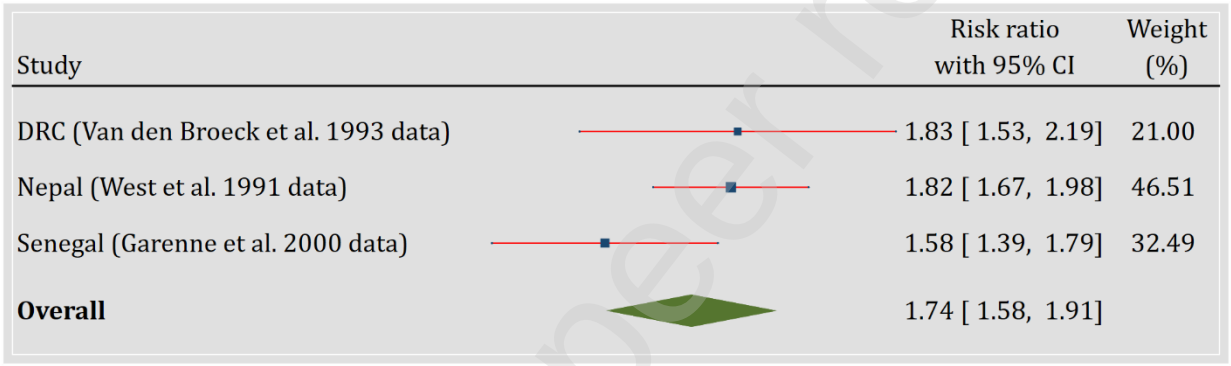


Table 11. Meta-analytical mortality ratios for malnourished vs. well-nourished children using 3-month follow-up, MIG model

	Burkina Faso			Chad			DRC			Ethiopia			Mali			Niger			Nigeria		
	WHZ low	WHZ high	Risk ratio	WHZ low	WHZ high	Risk ratio	WHZ low	WHZ high	Risk ratio	WHZ low	WHZ high	Risk ratio	WHZ low	WHZ high	Risk ratio	WHZ low	WHZ high	Risk ratio	WHZ low	WHZ high	Risk ratio
MAM	−2.41 (0.28)	−0.30 0.00	3.54 (1.03)	−2.42 (0.28)	−0.25 0.00	3.70 (1.12)	−2.44 (0.29)	−0.07 0.00	3.96 (1.25)	−2.31 (0.25)	−0.20 0.00	3.53 (1.02)	−2.40 (0.28)	−0.35 0.00	3.45 (0.99)	−2.42 (0.28)	−0.52 0.00	3.36 (0.94)	−2.39 (0.29)	−0.16 0.00	3.59 (1.06)
SAM	−3.69 (0.47)	−0.30 0.00	10.86 (5.46)	−3.62 (0.47)	−0.25 0.00	10.92 (5.57)	−3.72 (0.50)	−0.07 0.00	12.58 (6.76)	−3.53 (0.40)	−0.20 0.00	11.14 (5.77)	−3.70 (0.55)	−0.35 0.00	10.83 (5.44)	−3.71 (0.55)	−0.52 0.00	10.17 (5.00)	−3.57 (0.45)	−0.16 0.00	10.01 (4.92)

Note: WHZ values are means of DHS samples. Standard deviations (for WHZ low) and standard errors (for risk ratios) in parentheses. "WHZ high" values have zero standard deviation because the exact value shown is used for every subject in the samples.

Table 12. Meta-analytical mortality ratios for proxy treatment vs proxy control groups using 3-month follow-up, MIG model

	Burkina Faso			Chad			DRC			Ethiopia			Mali			Niger			Nigeria		
	WHZ low	WHZ high	Risk ratio	WHZ low	WHZ high	Risk ratio	WHZ low	WHZ high	Risk ratio	WHZ low	WHZ high	Risk ratio	WHZ low	WHZ high	Risk ratio	WHZ low	WHZ high	Risk ratio	WHZ low	WHZ high	Risk ratio
MAM, gov.	−2.41 (0.28)	−1.64 (0.28)	1.82 (0.23)	−2.42 (0.28)	−1.65 (0.28)	1.84 (0.24)	−2.44 (0.29)	−1.67 (0.29)	1.84 (0.24)	−2.31 (0.25)	−1.54 (0.25)	1.83 (0.24)	−2.40 (0.28)	−1.63 (0.28)	1.82 (0.24)	−2.42 (0.28)	−1.65 (0.28)	1.84 (0.24)	−2.39 (0.29)	−1.62 (0.29)	1.81 (0.23)
SAM, gov.	−3.69 (0.47)	−2.42 (0.47)	3.01 (0.63)	−3.62 (0.47)	−2.28 (0.47)	3.28 (0.75)	−3.72 (0.50)	−2.46 (0.50)	3.10 (0.66)	−3.53 (0.40)	−2.19 (0.40)	3.41 (0.81)	−3.70 (0.55)	−2.36 (0.55)	3.21 (0.71)	−3.71 (0.55)	−2.37 (0.55)	3.23 (0.72)	−3.57 (0.45)	−2.23 (0.45)	3.17 (0.71)
MAM, NGO	−2.41 (0.28)	−1.63 (0.28)	1.84 (0.24)	−2.42 (0.28)	−1.64 (0.28)	1.85 (0.25)	−2.44 (0.29)	−1.66 (0.29)	1.87 (0.25)	−2.31 (0.25)	−1.51 (0.25)	1.87 (0.26)	−2.40 (0.28)	−1.59 (0.28)	1.87 (0.26)	−2.42 (0.28)	−1.64 (0.28)	1.85 (0.25)	−2.39 (0.29)	−1.61 (0.29)	1.82 (0.24)
SAM, NGO	−3.69 (0.47)	−2.19 (0.47)	3.64 (0.90)	−3.62 (0.47)	−1.92 (0.47)	4.40 (1.27)	−3.72 (0.50)	−2.22 (0.50)	3.77 (0.95)	−3.53 (0.40)	−1.95 (0.40)	4.17 (1.17)	−3.70 (0.55)	−2.12 (0.55)	3.91 (1.02)	−3.71 (0.55)	−2.19 (0.55)	3.75 (0.94)	−3.57 (0.45)	−1.94 (0.45)	3.97 (1.08)

Note: WHZ values are means of DHS samples. Standard deviations (for WHZ low and WHZ high) and standard errors (for risk ratios) in parentheses.