General Equilibrium Effects of Cash Transfers: Pre-analysis plan for Endline 3 (EL3) Child Mortality Analysis[1]

Dennis Egger, Johannes Haushofer, Edward Miguel, and Michael Walker

May 10, 2023

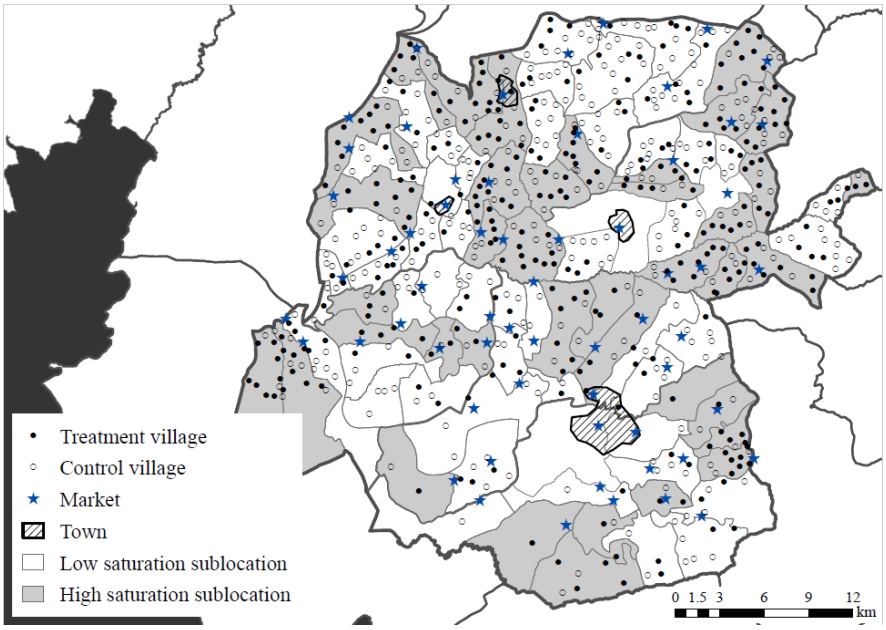
Summary: This document outlines outcomes and regression specifications for estimating the effects of unconditional cash transfers on child mortality as part of the General Equilibrium Effects (GE) project in western Kenya. The project is a two-level randomized controlled trial of the NGO *GiveDirectly*'s unconditional cash transfer program. Transfers were distributed from 2014-16, with a first set of findings described in Egger et al. (2022). As part of a third round of endline surveys, an additional household census will be carried out to identify births and child survival throughout the 653 villages within the study area. This document describes the analyses that will be conducted to understand the effects of the program on child mortality. We specify regression equations and the primary outcomes that we intend to study based on the household census data that will be collected as well as on existing data collected as part of previous rounds. This document is part of a series of pre-analysis plans that will be filed as part of the broader third endline (EL3) survey data collection activity. We anticipate conducting analyses beyond those pre-specified here, and this document is not meant to preclude additional analyses.

[1] AEA Trial Registry: AEARCTR-0000505, https://www.socialscienceregistry.org/trials/505. Egger: University of Oxford, dennis.egger@economics.ox.ac.uk; Haushofer: Stockholm University, johannes.haushofer@su.se; Miguel: University of California, Berkeley, emiguel@berkeley.edu; Walker: University of California, Berkeley, mwwalker@berkeley.edu. We thank Jasmin Baier, Grady Killeen and Layna Lowe for excellent research assistance in developing this pre-analysis plan. Funding from Open Philanthropy, recommended by GiveWell, is supporting this work.

1 Introduction

This document outlines the analysis plan for child mortality data collected between 2014 - 2023 (Baseline, Endline 1, Endline 2, Endline 3), as part of the General Equilibrium Effects (GE) project, a randomized controlled trial of an unconditional cash transfer program by the NGO GiveDirectly (GD). In villages selected for treatment, GD transferred around USD 1,000 (nominal) to all eligible households in the village, about 75% of annual expenditure for recipient households. Our transfers constituted a shock of about 15% of local GDP at the time that they were distributed. Only households with grass-thatched roofs were eligible to receive transfers, a basic means-test for poverty; we find 33% of households eligible in our study area. The intervention involved over USD 11 million in transfers and 653 villages in Siaya County, Western Kenya. Treatment assignment was randomized at the village level, and within treatment villages, all households meeting GD's eligibility requirement received the unconditional cash transfer. A second level of randomization provided variation in treatment intensity: sublocations, an administrative unit directly above the village including about ten villages on average, were randomly assigned to high or low saturation status. In high saturation sublocations, two-thirds of villages were treated, while in low saturation sublocations, only one-third of villages were treated. Figure 1 gives an overview of the study area and experimental design[2].

Figure 1: Study area and experimental design (from Egger et al. 2022)



There is currently little evidence for effects of unconditional cash transfers on child mortality (Abdul Latif Jameel Poverty Action Lab (J-PAL), 2020). This study is a unique opportunity to study such effects in a low or lower middle income setting both because (i) it uses experimental variation; (ii) it has a large sample size (over 65,000 households in the baseline census exercise), increasing statistical power to

[2] Additional details on the experimental design and implementation can be found in Egger et al. (2022) and Haushofer et al. (2017a).

detect effects of a relatively rare event (childhood death), and (iii) recipient households experience positive economic effects in the short-run (Egger et al. 2022), and preliminary results of analysis carried out to date from Endline 2 suggest some persistent economic effects 5 to 7 years later.

This analysis is primarily focused on data related to child mortality collected in Endline 3 (about 9 years after the transfer) although where appropriate, we will make use of earlier rounds of data as well. As such, this document draws heavily on previously filed pre-analysis plans and existing working papers, updating earlier regression specifications and outcomes. These earlier plans are described in section 1.1 below.

Broadly, there are two components to this analysis:

i. Studying the impacts of cash transfers on child mortality on recipient households (and their neighbors) present in the study area at the time of the intervention
ii. Identifying the impacts of cash transfers on the causes of child deaths using machine learning methods on verbal autopsy data

The primary data source for this analysis is a new round of household censuses in all study villages (the Endline 3 (EL3) Household Census). The EL3 household census data may be complemented with data to be collected from EL3 household surveys (from a representative sample of households), as well as data collected as part of previous survey rounds (in particular EL2). The representative household sample includes households that were residents in the study area at baseline, as well as some households that moved into the study area after the intervention ('new households'). We carefully track and survey both households that have remained in the area and those that have moved away, either since baseline or afterwards (i.e., if they were added into the sample as 'new households' but have since moved away again). We describe the sample to be used to measure the effects of cash transfers on child mortality in detail further below (see Section '5.2 Econometric specification of impact analysis').

This is the first pre-analysis plan that we are filing as part of EL3 data collection activities. In addition to this pre-analysis plan, we currently plan to file (i) a pre-analysis plan on the enterprise census activity; (ii) a pre-analysis plan on migration outcomes; and (iii) pre-analysis plans covering the household and enterprise surveys, once these instruments are finalized, in order to study the longer-term effects on welfare of recipient households and local economy effects (in line with Egger et al. 2021a and Egger et al. 2021b from Endline 2). There may also be additional analyses based on these data in the future.

Endline 3 census data collection began on April 20th, 2023. We are filing this PAP shortly after the launch, with about 10% of expected observations. To date, no member of the research team has linked the data to treatment indicators nor estimated any treatment effects. After filing this plan, we will begin estimating pre-registered effects for household outcomes.

1.1 Relation to previous work

This pre-analysis plan builds on a series of earlier pre-analysis plans filed for the GE project as part of data collection in previous rounds (short-term/Endline 1: (Haushofer et al. 2016, Haushofer et al. 2017a, Haushofer et al. 2017b, Haushofer et al. 2018, Walker 2017); Endline 2: Egger et al. (2021a, b), Orkin and Walker 2021, Egger et al 2022). Moreover, it builds on analyses published in Egger et al. (2022), as well

as ongoing analyses of Endline 2 data, which suggests there is some persistence in economic gains for household consumption and assets, among other effects. In this PAP, we err on the side of brevity to avoid repetition, as much of the detail and thought development can be found in these earlier documents.

2 Research Design

2.1 Sampling and treatment assignment

The GE project takes place in Siaya County, Kenya, a rural area in western Kenya bordering Lake Victoria. Siaya County is predominantly Luo, the second largest ethnic group in Kenya. GD selected both Siaya County and a region within Siaya County based on its high poverty levels and identified target villages for expansion; in practice, these were all villages within the region that a) were not located in peri-urban areas and b) were not part of a previous GD campaign. This gives a final sample of 653 villages, spread across 84 administrative sublocations (the unit above a village), and 3 constituencies.

We use a two-level randomization in order to generate variation that can be used to identify spillover effects. We randomly assigned sublocations (or in some cases, groups of sublocations) to high or low saturation status. Then, within high saturation groups, we assigned 2/3 of villages to treatment status, while within low saturation groups, we assigned 1/3 of villages to treatment status. As noted above, within treatment villages, all eligible households received a cash transfer.

At baseline, we censused about 65,000 households. For this round, we anticipate similar (or slightly higher) numbers. A random sample of households were then drawn for detailed household surveys at baseline; these households were followed up at Endline 1 and Endline 2. Additionally, at Endline 2, we added a random sample of new households to the sampling frame. We anticipate taking a similar approach for household surveys at Endline 3, though sampling plans for household surveys will be finalized once the household census is complete.

2.2 Intervention

GD provides unconditional cash transfers to poor households in rural Kenya, targeting (for villages in the study) households living in homes with thatched roofs, a basic means-test for poverty. In treatment villages, GD enrolls all households in treatment villages meeting its thatched-roof eligibility criteria (“eligible" households); approximately one-third of all households are eligible. No households in control villages receive transfers. Eligible households enrolled in GD's program receive a series of 3 transfers totaling about USD 1,000[3] via the mobile money system M-Pesa. This is a one-time program and no additional financial assistance is provided to these households after their final large transfer. For details on the intervention, see Egger el al. (2022) and Haushofer et al. (2017a).

[3] The total transfer amount is 87,000 Kenyan Shillings (KES). The average exchange rate from 9/1/14 to 4/30/16 was 97 KES/USD.

2.3 Data and Instrument

The analyses outlined in this document will be primarily based on a new round of data collected in 2023 (Endline 3), roughly 8-9 years after the GD cash transfers went out, as highlighted in the approximate timeline below:

- 2014 - 2015: Baseline (pre-intervention)
- 2014 - 2016: Intervention
- 2014 - 2017: Midline
- 2016 - 2017: Endline 1
- 2019 - 2022: Endline 2
- 2023 - 2024: Endline 3

Activities for Endline 3 will include a household census, enterprise census, household surveys, and enterprise surveys. The household and enterprise censusing activities will be conducted at the same time, and are estimated to last from approximately April 2023 to September 2023. The survey activities will also be conducted concurrently with each other, beginning after the conclusion of the census.

The child mortality analysis we pre-specify here will primarily make use of data from the household census in our study area, although as mentioned in other parts of this PAP, we will also bring in information from EL2 and EL3 household surveys. Specifically, EL3 household surveys will be important to fill in the child mortality and verbal autopsy information of those households that moved away from the study area and as such will not be surveyed in the household census. (A representative subsample of households will be surveyed in the EL3 household survey, including those who were present in earlier rounds but no longer live in the study area.) The household census involves re-visiting each 653 villages, and working with local leaders and past census / tracking information to a) identify each household within the study village and b) collect a short census module with a member of each household. The household census will identify if households have remained in the study area, as well as counting newly established households. For each household, we also collect a roster of births that have occurred since 2011 for female household members that have lived in the household for at least 4 months.[4] For each birth, we then collect information on the date of birth, birth place, as well as potential migration into and out of the study zone. We also collect information on the current status of the child, namely whether or not they are living or have died. If no longer living, we collect the date of death, age at death, and the verbal autopsy module (described further below).

There are two scenarios of births that we will miss under this approach. In initial pilot work in the study area, we believe these cases are sufficiently rare so as to not meaningfully alter the analysis:

1. Those cases where the mother never lived in the study area
    a. Because the birth history module starts with asking about women who have lived in this household, we miss births from women who never lived in the study area but whose children may have lived in the study area for at least some time since 2011.

[4] We consider all births in which the child could have had some exposure to the intervention (occurred in 2014-5) or its longer-term effects,, at least within the first five years of his/her life.

   b. Similarly, we may miss children whose mother did not live in the study area but who lives/lived with a single father, grandparent, other relative, etc.
2. Those where the mother left the study area, and the household within the study area that she was a member of no longer exists
   a. If the only household that a woman had membership in no longer exists, for instance, because all members are deceased or have moved away, then we will miss her birth history.

Once the number of child deaths under 5 in a household in or after 2011 has been identified, the WHO 2022 Verbal Autopsy Questionnaire will be administered for each child. Verbal autopsy (VA) is a method of determining individuals' causes of death and cause-specific mortality fractions in areas lacking the infrastructure to perform physical autopsies on all the deceased and store that data in a vital registration system, as is standard in most high income countries today. Verbal autopsies usually consist of an enumerator using a questionnaire to collect information about a deceased person's signs, symptoms, and demographic characteristics from an individual familiar with the deceased. The WHO provides clear guidance about who is the best person to respond to these questions within a household.[5] In the case of deceased in the 0-6 age group, that will most likely be their mother or primary caregiver at the time of death.

VA is the state-of-the-art survey-based method for determining causes of death based on self-reported information. Both the data collection method as well as cause identification methods from verbal autopsy data have been extensively validated. Serina et al. (2015), for example, find that when comparing the VA's performance to the true causes of so-called "gold standard deaths" as defined by the Population Health Metrics Research Consortium (PHMRC), the Tariff 2.0 algorithm with healthcare experience (HCE) coding displayed a 78.3% median accuracy for child deaths, and 82.8% for neonate deaths across 500 iterations of calculating the cause-specific mortality fractions (CSMF).

The tool has been extensively used in the research context of this project and even in the specific study area (Gacheri et al, 2004; Nyaguara et al., 2014; Amek et al., 2018), though we recognize the limitations mentioned in these studies, such as a lower accuracy in exact cause of death identification. In the interpretation of the findings that result from the data, we will be mindful of such potential limitations, for example by analyzing broader cause of death groups. Moreover, we recognize the limitation of the long recall period (back to 2011), which has not been rigorously assessed in previous work. Most studies have a much shorter recall period, especially when used for surveillance, and we are not certain about how a recall period of up to 12 years may affect the ability of the algorithm to determine a single probable cause of death.

The research team behind this pre-analysis plan has been working closely with, and learning from the expertise of, the local Kenya Medical Research Institute (KEMRI) research team that has been implementing a Health and Demographic Surveillance System (HDSS) located in Rarieda, Siaya and Gem Districts (Siaya County) since 2001 (Odhiambo et al., 2012). This exercise includes verbal autopsies

[5] We anticipate using 2011 as the default end of our recall period, but we may adapt this rule if we find that respondents have challenges with recall in the early years of this period.

of all deaths in the study area. Both enumerator training and data analysis will be closely coordinated with existing relationships with KEMRI staff.

4 Outcomes of interest

The main measure of interest focuses on child mortality. In addition, the VA tool employed in this study primarily leads to two distinct outcomes: a determined cause of death for each individual and the cause-specific mortality fraction (CSMF), and we will also use these two measures.

The following outcomes will primarily stem from Endline 3 household census data, although we plan to to integrate this with some aspects of previous data (including a more limited set of child mortality measures available in Endline 2 data) and data from future data collection (Endline 3 household surveys).[6]

4.1 Primary outcome of interest

4.1.1 **Child mortality:** Birth-level variable that takes the value 1 if the child died before turning 5, and 0 otherwise. The population to be analyzed in this measure consists of children who were born at least 5 years before the time of data collection in order to have a consistent population for both the numerator (children who have died) and the denominator (children who have died plus children who have survived until the age of 5). Observations of children who are alive but below 5 years old are not included. Similarly, children who have died but were born within 5 years before data collection are not included.

4.1.2 **Infant Mortality:** Similar to 4.1.1, but focusing the analysis on children who died at age 0-1, where we again include all children whose birth date was at least 1 year before the time of data collection, and we exclude all children whose birth dates were less than a year before data collection (regardless of whether or not they survived).

4.2 Secondary outcomes of interest

We have two specific families of secondary outcomes of interest: (i) additional child mortality and survival outcomes, and (ii) cause of death outcomes.

Family 1: Additional child mortality and survival outcomes

4.2.1 **Neonatal Mortality:** Similar to the mortality metrics used as primary outcomes, but focusing the analysis on children who died within 0 and 28 days of birth, where we again do not include children whose birth date was within 28 days before the time of data collection. We consider this outcome as secondary because the literature suggests that neonatal mortality is more closely linked to congenital causes, rather than household socioeconomic circumstances or access to medical care, and thus it is less

[6] The child mortality measure we can construct with Endline 2 survey data only includes children of the respondent or the respondent's spouse, rather than all female household members, as we attempt to capture in Endline 3. This may potentially help resolve some of the missing cases noted above.

closely linked to the program we study than the primary outcomes of under-5 mortality and infant mortality.

4.2.2 **Days of survival:** The number of days a child survived (up to 5 years of age). This continuous measure is not widely used (to our knowledge) in the public health literature but has some desirable properties as it combines aspects of neonatal, infant and child mortality and avoids arbitrary cutoffs. It provides another way to capture the intensive margin of child survival; for instance, we would capture the survival gains from a child who lived to 4 years and 364 days versus 1 year and 1 day while these would not be captured in the child mortality and infant mortality indicators alone. It is also potentially useful for value of a statistical life year (VSLY) calculations.

Family 2: Cause of death outcomes

4.2.3 **Cause of death:** Absolute number and percentage shares of specific causes of death. We will perform both analysis *on each individual cause of death (birth-level dummy variable)*, as well as a *group-level analysis*. For the group-level analysis, we will first use the standard three-group classification built into most VA algorithms plus an "undetermined" category:

4.2.3.1a, *Communicable, maternal, neonatal and nutritional diseases* - including the individual causes: AIDS, Diarrhea/Dysentery, Encephalitis, Hemorrhagic fever, Malaria, Measles, Meningitis, Other Infectious Diseases, Sepsis, Neonatal Meningitis/Sepsis, Birth asphyxia, Congenital malformation, Preterm Delivery, Stillbirth, Pneumonia, Neonatal Pneumonia, and tuberculosis related disease

4.2.3.2a, *Injuries* - including the individual causes: Bite of Venomous Animal, Drowning, Falls, Fires, Poisonings, Road Traffic, Homicide (assault)

4.2.3.3a, *Non-communicable diseases* - including the individual causes: Child Cancers, Child Cardiovascular Diseases, Other Defined Causes of Child Deaths, Other Digestive Diseases

4.2.3.4a *Undetermined causes*

In a secondary examination, we group causes into six distinct groups in order to get slightly deeper insights:

4.2.3.1b*, Nutritional disorders, and non-respiratory communicable diseases* - including the individual causes: AIDS, Diarrhea/Dysentery, Encephalitis, Hemorrhagic fever, Malaria, Measles, Meningitis, Other Infectious Diseases, Sepsis, Neonatal Meningitis/Sepsis

4.2.3.2b *Maternal and specifically neonate causes* - including the individual causes: Birth asphyxia, Congenital malformation, Preterm Delivery, Stillbirth

4.2.3.3b *Respiratory diseases* - including the individual causes: Pneumonia, Neonatal Pneumonia, and tuberculosis related disease

4.2.3.4b *Injuries* - including the individual causes: Bite of Venomous Animal, Drowning, Falls, Fires, Poisonings, Road Traffic, Homicide (assault)

4.2.3.5b *Non-communicable diseases (NCD)* - including the individual causes: Child Cancers, Child Cardiovascular Diseases, Other Defined Causes of Child Deaths, Other Digestive Diseases

4.2.3.6b *Undetermined causes*

More details on cause of death identification is described in section 5.1.

5 Analysis

The analysis of data collected that is relevant for this PAP can be separated into two stages: moving from raw VA data to determined causes of death, and subsequently using the constructed mortality variables to produce impact estimates. Both processes will be described in detail in the next two sections.

5.1 Cause of death identification

The most commonly used approach to analyzing VA data has been physician review. However, this method is costly due to labor costs and the lengthy process involved. Increasingly, it is also shown to be less effective than established machine learning algorithms designed to be used alongside VA data to establish the most likely causes of death for each person being autopsied. Byass et al. (2015), for example, find that one of the algorithms, InterVA, shows significant correspondence with physician assignment, and as such can be said to produce fairly reliable cause of death assignment even across a diverse dataset. The introduction of ML thereby not only reduces costs but greatly improves the reproducibility and open science credentials of VA.

According to the WHO, there are four different algorithms leading to reliable cause of death diagnoses. They do not provide a clear recommendation among them but instead leave it to the research team to find the most suitable approach. The four algorithms, in short, are:

1. **InterVA:** The reported symptoms are combined with probabilities (produced by medical experts) of the likelihood for the observed symptoms in the case of a specific cause of death. The highest propensity determines the cause of death unless none of the scores reach above a certain threshold (in which case it would be "indeterminate"). The cause-specific mortality fraction (CSMF) is the percentage of each cause of death out of all deaths as found in the data itself.
2. **InSilicoVA:** Both the presence and absence of each symptom are considered in calculating probabilities. Moreover, the CSMF is not purely calculated from the data but includes a pre-trained statistical model.
3. **Naïve Bayes Classifier**: Similar to InSilicoVA, this algorithm considers both the presence and absence of symptoms, but, similar to InterVA, uses the distribution of assigned causes in the data to estimate the CSMF.
4. **Tariff and Tariff 2.0:** "Tariffs" estimate how informative each symptom is for a specific cause of death, based on the PHMRC gold standard VA validation study dataset (Murray et al., 2011). The sum of Tariffs is then matched to the cause of death with the highest ranked Tariff Score. Tariff 2.0 is an update to this approach that improves reliability by having removed or adjusted symptom-cause associations from the initial model. The CSMF is calculated from the distribution in the data, as in InterVA.

More information on the mathematical properties of these algorithms can be found here.

As of now, most of these algorithms have not been updated to the 2022 version of the WHO's VA instrument. Furthermore, it does not seem feasible to restructure data from the 2022 version of the instrument to resemble the 2016 version enough for the algorithms to function properly. This is because

the 2022 instrument is much shorter, and therefore not only asks fewer questions but also asks questions slightly differently to capture more information in a shorter interview time.

The analysis program which has already been updated, SmartVA, uses the Tariff 2.0 algorithm and was used to analyze pilot data we have collected up to now. The researchers may choose a different algorithm for the main analysis if other programs have been updated by that point. In particular, the researchers will prioritize algorithms which seem theoretically most suitable and/or those used by other organizations and researchers in the same context. It currently seems likely that the InterVA algorithm, which is also being used by the Kenya Medical Research Institute (KEMRI), is most suitable for the purposes of this study, though this perception may change as research continues to evolve in this space. It is furthermore unclear as of now if that algorithm will be updated to function for the WHO 2022 instrument. The KEMRI research team has also recently estimated infant and child causes of death in the study region using a similar methodology, which may be useful to discipline priors and serve as a cross-check for our cause of death analysis.

5.2 Econometric specification of impact analysis

The main purpose of this analysis plan is to study the long-term effects of cash transfers on child mortality. The data allow for us to look at several different effects related to this, depending on the sample that we use for estimating effects.

First, we look at households that were present at Baseline, and that remain present at Endline 3. These households were present at the time of the cash transfer, and thus have a clearly defined treatment and eligibility status; the fact that they remain present at Endline 3 ensures that we have full child mortality data from them from the EL3 household census. We include Baseline Households that have stayed in the same location, as well as Split Households (i.e. those that have a ‘parent’ household that was in the study area at Baseline but have established a new household for themselves; for these treatment and eligibility status is assigned based on the ‘parent’ household).

This analysis misses households which were present at baseline but later moved away; as this is an endogenous decision that may have implications for child survival, it is interesting to understand how estimates may change when taking migration out of the study area by households present at the time of the transfers into account. As part of Endline 2 household surveys, conducted with a random sample of the full study area population, we (i) tracked households that moved out of the study area and (ii) collected basic child mortality data on children of the focus respondent (not the full household), which we can use to generate some initial insights around child mortality for migrants. We plan to continue tracking migrants as part of Endline 3 household surveys, and will collect an analogous child mortality module to the one included as part of the earlier EL2 household census. As household survey respondents are a representative sample of households living in the study area at Baseline, we calculate child mortality effects for all households in the study area at baseline by appropriately combining and weighting household census and survey estimates using inverse sampling probability weights.

One may also be interested in whether the intervention affected child mortality for the study area as a whole at endline, for which we would want to bring in New Households (those established in the study

area after cash transfers went out), somewhat analogous to Endline 2 local economy estimation (Egger et al. 2021b).

Primary specifications (i.e. (1), (2), and (3)) only include births after 2014 (i.e. after cash transfer treatment began). Births from before will only be considered in secondary analyses (see '5.2.3 Effects over time'). In the case that we find a changing effect over time (see '5.2.3 Effects over time'), we may explore alternative definitions of time periods (e.g. pool the first 3 to 5 years after treatment for more power). When calculating the mortality rate, we will follow the same approach as UNICEF and only include children who were born at least 5 years before the time of data collection, as described in section '4 Outcomes of interest'. As an example, in our primary analysis, children who are or would have been e.g. three years old at the time of data collection will be excluded from the child mortality metric, and children who are or would have been under age 1 will be excluded from the infant mortality metric. We may choose to include the whole sample for a robustness check as an additional approach.

The econometric analysis largely follows Egger et al. (2022). First, we estimate the effects on recipient households, then the effects on non-recipient households present in the study area at the time of transfers (Baseline). **We consider effects on eligible recipient households (in equations 1 and 2 below) to be the primary effects of interest.** Second, we also plan to present results for non-recipient households (i.e., both eligible non-recipients and all ineligibles) to capture spillover effects (as in equation 3 below). Third, we plan to estimate pooled effects among both recipients and non-recipients as an additional specification (as in equation 4 below); while the pooled sample may be better powered statistically given its larger sample size of households and births, it is also not directly comparable to the analysis among recipient households since non-recipients did not experience any direct cash transfers (and thus may also have experienced smaller economic gains). That said, the pooled specification has the advantage of capturing the average impact of the cash transfer program that we study on local infant and child survival, and thus may be of interest as part of the overall program impact evaluation.

5.2.1 Effects on recipient households

Recipient households experience both direct effects of cash as well as potential within and across-village spillovers. A useful benchmark is the following specification which we run for eligible households present in the study area at the time of transfers:

$$y_{imhvs} = \alpha_1 Treat_v + \alpha_2 HighSat_s + \lambda_{t(i)} + \rho_{g(i)} + \lambda_{t(i)} * \rho_{g(i)} + A_m + \delta M + \varepsilon_{imhvs} \quad (1)$$

Where y_{imhvs} is an outcome for a birth i, in household h, in village v and sublocation s (at Baseline), $Treat_v$ is an indicator for village v being treated, and $HighSat_s$ is an indicator for sublocation s being allocated high saturation status (i.e. 2⁄3 of villages being treated in s, as opposed to 1⁄3 in low-saturation sublocations). $\lambda_{t(i)}$ and $\rho_{g(i)}$ denote the child's year of birth and gender fixed effects. A_m stands for dummy variables indicating the mother's (m) age in five age groups (under 20, 20 to 25, 25 to 30, 30 to 35, above 35). **M** is a vector of missingness indicators for each of the covariates which allows us to retain observations in order to maximize power.

We cluster standard errors at the village level, and weight observations by inverse sampling probabilities to be representative of the population of eligible households.

The primary estimation approach will use a linear probability model (OLS), but as a robustness check, we will consider logit and probit models. Moreover, we will likely run this and any of the following specifications also without any fixed effects.

The primary parameter of interest is α_1 which captures the direct and within-village spillover effects of treatment on the recipients' households' births/children within treatment villages. α_2 measures any additional across-village spillovers within sublocations (but not across). Since administrative sublocation boundaries in the study's context are relatively unimportant for economic activity, with substantial interactions across borders, and no meaningful social or ethnic divides, we do not consider α_2 primary. Instead, the main spillover measure will be captured in equation (2).

To better capture the full spatial dimension of spillovers, we estimate the following specification for eligible households present in the study area at Baseline:

$$y_{imhvs} = \beta Amt_v + \sum_{r=2}^{\overline{R}} \beta Amt_{v,r}^{\neg v} + \lambda_{t(i)} + \rho_{g(i)} + \lambda_{t(i)} * \rho_{g(i)} + A_m + \delta M + \varepsilon_{imhvs} \quad (2)$$

where, following Egger et al (2022), Amt_v is the total per-capita amount transferred to village v, $Amt_{v,r}^{\neg v}$ is the total per-capita amount transferred to households in a buffer of r-2 to r km around the village centroid of village v, and all other variables are defined as above. The *Amt* variables depend on both the random assignment of villages to treatment and also on the endogenous share of households in those villages eligible for transfers, so we instrument for them using the own-village treatment indicator *Treat* and the share $s_{\neg v,r}^{e,t}$ of eligible households in each band assigned to treatment (at the time of treatment), again as in Egger et al (2022). All other variables defined as above. To account for spatial correlation, we calculate standard errors using a uniform kernel up to 10 km (Conley 2008).

We select the maximum radii band included in the main specification ($\overline{R}$) as in Egger et al. (2022), by first estimating a series of nested models with the outer limit R varying from 2 to 20 km, and then selecting the model which minimized the Schwarz Bayesian Information Criterion (BIC). In addition to the maximum radius selected by this procedure, we will also analyze effects using $\overline{R}_{SR}$, where $\overline{R}_{SR}$ is the 'optimal' radius selected in the short run, for most outcomes at 2 km with a few exceptions at 4 km (Egger et al. 2022). If effects remain strongly localized as in Egger et al. (2022), we may additionally explore the spatial structure of spillovers, by defining smaller increments between radii bands.

The main parameter of interest is the average "total effect" (including direct, within-village and across-village spillovers) experienced by recipient households, as in Egger et al. (2022). We calculate

these by multiplying the estimated coefficients of Equation (2) by the average values of the regressors, i.e.

$$\widehat{\Delta y^{r}} = \widehat{\beta} * (Amt_{v} \mid i \text{ is an eligible recipient}) + \sum_{r=2}^{\overline{R}} * (Amt_{v,r}^{\neg v} \mid i \text{ is an eligible recipient}).$$

5.2.2 Effects on non-recipient households

Non-recipients are composed of eligible households in control villages as well as ineligible households in both treatment and control villages. These households experience indirect effects of cash transfers through both within-village spillovers (ineligibles in treated villages only) and across-village spillovers (both ineligibles and eligibles in control villages). Here, the aim is to compare the effects on recipients to non-recipients living in the study area at Baseline. We estimate:

$$\begin{aligned} y_{imhvs} = & \sum_{r=2}^{\overline{R}} \beta_r^1 Amt_{v,r} + \sum_{r=2}^{\overline{R}} \beta_r^2 (Amt_{v,r} * Elig_{h,v}) + \gamma Elig_{h,v} \\ & + \lambda_{t(i)} + \rho_{g(i)} + \lambda_{t(i)} * \rho_{g(i)} + A_m + \delta M \\ & + \lambda_{t(i)} Elig_i + \rho_{g(i)} Elig_i + \lambda_{t(i)} * \rho_{g(i)} * Elig_i + A_m * Elig_i + \delta M * Elig_i + \varepsilon_{imhvs} \end{aligned} \quad (3)$$

Where $Amt_{v,r}$ is the per-capita amount of cash transferred into the r-2 to r km buffer around the centroid of village v, and $Elig_{h,v}$ is an indicator for household h being eligible, non-eligible, or not in the area at baseline. All other variables defined as in Equation (1). We again instrument for $Amt_{v,r}$ using the share of eligibles within buffer r around village v allocated to treatment ($s_{v,r}^{e,t}$), and use Conley (2008) standard errors with a uniform Kernel up to 10 km. The procedure for determining $\overline{R}$ will be the same as noted above for Equation (2).

The main parameter of interest is the 'total effect' on non-recipient households, which is a population-weighted average of the average effects experienced by eligible households in control villages and ineligible households, as in Egger et al. (2022). This is calculated as:

$$\widehat{\Delta y^{n}} = s^{e,c} * (\sum_{r=2}^{\overline{R}} (\widehat{\beta}_r^1 + \widehat{\beta}_r^2) * \overline{Amt}_{v,r} \mid i \text{ is an eligible non } - \text{ recipient})$$

$$+ s^{i} * (\sum_{r=2}^{\overline{R}} \widehat{\beta}_r^1 * \overline{Amt}_{v,r} \mid i \text{ is ineligible})$$

Where $s^{e,c} = 1 - s^{i}$ is the population share of eligible non-recipients among all non-recipient households. In supplemental material, we report the effects for these two groups separately, as well as the breakdown between within- and across-village spillovers (see the Appendix of Egger et al. (2022)).

The main specifications (1), (2) and (3) do not include any covariate adjustments, since Egger et al. (2022) found these to leave main results largely unchanged. We may consider additional specifications

with covariate adjustments as described in Haushofer et al. (2017a) if this meaningfully increases the precision of the estimates.

5.2.3 Pooled effects on all households

To calculate the pooled effect on all households living in the study area at baseline – i.e. both recipients and non-recipients, we will estimate a population-weighted average treatment effect on these two groups of households, following Sections 5.2.1 and 5.2.2 from Equations (2) and (3) respectively. This is calculated as:

$$\widehat{\Delta y} = \frac{N_r}{N_r+N_n}\widehat{\Delta y^r} + \frac{N_n}{N_r+N_n} * \widehat{\Delta y^n} \qquad (4)$$

where N_r and N_n are the population number of recipients and non-recipients respectively.

5.2.4 Effects over time

We are interested in two primary time effects: (i) effects of cash over time, and (ii) heterogeneity of effects depending on timing and intensity of the intervention within a child's first five life years. We would like to study these with the following specifications:

(i) Does the effect of cash on child mortality change over time?

To calculate this effect, we first estimate the dynamic equations documented in Egger et al. (2021a):

$$y_{imhvst} = \sum_{l \in L} \beta_l Amt_{v(t-l)} + \sum_{l \in L}\sum_{r=2}^{\bar{R}} \gamma_{l,r} Amt^{\neg v}_{v(t-l),r} + \lambda_{t(i)} + \rho_{g(i)} + \lambda_{t(i)} * \rho_{g(i)} + A_m + \delta M + \varepsilon_{imhvs} \quad (5)$$

$$\begin{aligned} y_{imhvst} = {} & \sum_{l \in L} \beta^1_l Amt_{v(t-l),r} + \sum_{l \in L}\sum_{r=2}^{\bar{R}} \beta^2_{l,r}(Amt_{v(t-l),r} \cdot Elig_{iv}) + \gamma Elig_{iv} \\ & + \lambda_{t(i)} + \rho_{g(i)} + \lambda_{t(i)} * \rho_{g(i)} + A_m + \delta M \\ & + \lambda_{t(i)} Elig_i + \rho_{g(i)} Elig_i + \lambda_{t(i)} * \rho_{g(i)} * Elig_i + A_m * Elig_i + \delta M * Elig_i + \varepsilon_{imhvsw} \end{aligned} \quad (6)$$

Where y may be one of five outcomes: (i) all child deaths (up to age 5), (ii) neonatal deaths, (iii) under-1 deaths, (iv) deaths between 1 and 3 years, or (v) deaths between 3-5 years.

Subscript *t* refers to the year of birth of an individual. The data consists of a panel constructed via recall. For instance, if a mother reported a birth in 2018 and a birth in 2020, there would be two observations. The birth year fixed effects, $\lambda_{t(i)}$, function similarly to survey time fixed effects, but they capture time trends at the time the outcome occurred (birth time), rather than when it was recalled (survey time).

$Amt_{v(t-l)}$, $Amt^{\neg v}_{v(t-l),r}$, and $Amt_{v(t-l),r}$ are transfer amounts flowing into village *v* at time *t-l,* and into radii bands from r to r-2 from village *v* at time *t-l* (excluding or including village *v*). $\bar{R}$ is determined as

described above. $\gamma_{t(i)}$ are years-of-birth-fixed-effects. L are sets of time periods before time of birth, and $\gamma_{l,r}$ captures the effect of a marginal increase in the amount transferred *l* periods before one's birth.

A limitation of this econometric approach is that observations that were exposed to cash early on end up entering the control group for estimated effects on those that were exposed to cash late, a similar problem to those recently identified in the staggered difference-in-difference literature (e.g. Goodman-Bacon 2021). We are aware that this could attenuate dynamic estimates in the presence of persistent treatment effects, so we may consider alternative estimates based on developments in the econometrics literature (e.g. excluding observations that did not receive cash in *(t-l)* but were exposed to cash in an earlier period), including any new methods developed after the filing of this plan, as this is an active research area in applied econometrics.

We plan to use a yearly frequency as our default, but may explore other periodicities. $\rho_{g(i)}$ denotes the child's gender fixed effects. As in equation (1), A_m stands for dummy variables indicating the mother's age in five age groups (under 20, 20 to 25, 25 to 30, 30 to 35, above 35), and finally **M** is again a vector of missingness indicators for any of the covariates in order to maximize power.

We follow similar thinking to what is outlined in Egger (2021b, Section 6. Aggregate outcomes of interest) except that measurement of the relevant outcomes was not completed in real-time over several endline data collections, but will be based on an unbalanced panel through recall from Endline 3 data. Specification (6) attempts to measure spillovers similar to (3). We will estimate and assess the statistical significance of effects at different time lags relative to the distribution of cash transfers, along the lines of an impulse-response analysis. We will further examine the joint significance of the coefficients on the lagged terms, and estimate the cumulative effect of cash transfers using an approach related to the transfer multiplier estimates in Egger et al (2022).

(ii) Do effects of cash vary depending on what time of your life you are exposed (how many years and which ones)?

$$y_{imhvs} = \sum_{g \in G} \beta_g Amt_{vg} + \sum_{g \in G} \sum_{r=2}^{\bar{R}} \gamma_{g,r} Amt_{vg,r}^{\neg v} + \lambda_{t(i)} + \rho_{g(i)} + \lambda_{t(i)} * \rho_{g(i)} + A_m + \delta M + \varepsilon_{imhvs} \quad (7)$$

$$\begin{aligned} y_{imhvsw} = & \sum_{g \in G} \beta_l^1 Amt_{vg,r} + \sum_{g \in G} \sum_{r=2}^{\bar{R}} \beta_{l,r}^2 (Amt_{vg,r} \cdot Elig_{iv}) + \gamma Elig_{iv} \\ & + \lambda_{t(i)} + \rho_{g(i)} + \lambda_{t(i)} * \rho_{g(i)} + A_m + \delta M \\ & + \lambda_{t(i)} Elig_i + \rho_{g(i)} Elig_i + \lambda_{t(i)} * \rho_{g(i)} * Elig_i + A_m * Elig_i + \delta M * Elig_i + \varepsilon_{imhvs} \end{aligned} \quad (8)$$

Which considers the same outcomes as (5) and (6) as well as the same demographic fixed effects. $\lambda_{t(i)}$ remain year-of-birth fixed effects, which are also interacted with gender. Moreover, the amounts transferred now relate to the relevant groups, rather than time. The age groups g are defined as follows:

- Amount transferred between three and five years before birth

- Amount transferred between 9 months and three years before birth
- Amount transferred within 9 months before birth (treatment in-utero)
- Amount transferred up to 28 days after birth (treatment as neonate)
- Amount transferred between 28 days and one year before birth (treatment as infant)
- Amount transferred between 1 and 2 years after birth (treatment as young child)
- Amount transferred between 3 and 5 years after birth (treatment as older child)

Intuitively, this specification examines whether there are different effects on children that were exposed to cash in different times of their life, in utero, and after birth. Any effects of pre-birth exposure are more likely to work through effect on mothers. This addresses the mechanisms through which cash affects child mortality, whereas the first dynamic specification (in equations 5 and 6) estimates an impulse response function that tests whether the effects of cash diminish or grow over time. The main econometric difference between these two specifications is that we consider time gaps that are non-standard in equations 7 and 8 and concentrated near the time that cash was distributed, motivated by key stages of pregnancy. In contrast, equations 5 and 6 examine uniform time lags, and they focus on longer time gaps between the transfers and child birth year.

Since equations 7 and 8 examine the time of one's life when cash is most impactful, we will perform tests for equality across the coefficients on the average total effects for different groups *g,* in addition to looking at effects for particular periods *g*.

As with the specification in equations 4 and 5, using births where the individual was exposed to treatment at a different time in their life as a control could potentially bias estimates. We may consider estimates from the staggered difference-in-difference literature that are robust to these concerns.

A secondary interest relates to potential intergenerational effects, i.e. a child is treated around the ages 10-18 and has given birth to their own child since treatment. We are not confident that we will have the power to detect such effects at this time, but may run some exploratory analysis and re-consider this idea in subsequent data collection rounds.

As robustness check, we will also explore general mortality trends, i.e. distributions across age groups and death causes before and after the intervention in 2015.

5.2.4 Mechanisms

We anticipate that estimating effects on the cause of child deaths will provide some indication of potential mechanisms that may be behind any child mortality effects that we observe. In addition, we plan to use the census data plus the household survey data to explore other hypothesized channels for child mortality effects. For instance, we have detailed information on how treatment affected variables such as expenditure on health, nutrition, fertility, and other factors that are likely mechanisms for lower child mortality. We can also characterize correlations between these variables and child mortality in the data. This exercise may also be based on the heterogeneity analysis as described below (and those previously pre-specified as part of earlier plans), clustering algorithms using household characteristics, or panel data analysis.

5.4 Multiple inference adjustments

We have two primary outcomes – child and infant mortality – that are likely to be highly correlated. To account for multiple inference for our primary outcomes, we make use of the Romano-Wolf multiple testing correction, which asymptotically controls the Family-wise Error Rate (FWER). We plan to use a resampling based approach, rather than an analytic approach to limit the FWER, because it is able to mimic the dependence structure between child and infant mortality that we expect to see in the data, yielding better power.

For secondary outcomes, we follow Haushofer et al. (2017a) for our treatment of multiple inference adjustment, namely calculating sharpened q-values (i) across primary and secondary outcomes (as defined in section 4) following Benjamini, Krieger, and Yekutieli (2006) to control the false discovery rate (FDR) within each family of secondary outcomes. The FDR controls for the proportion of false positives, which is relevant if one is interested in the proportion of all outcomes affected by treatment. Rather than specifying a single q, we report the minimum q-value at which each hypothesis is rejected, following Anderson (2008). We will report both standard p-values and minimum q-values. We will apply the correction separately for each hypothesis test described in Section 3.1. We note that norms around multiple testing are still evolving in economics, and through the above methods seek to follow current best practices.

5.5 Balance and attrition

Egger et al. (2022) contains a host of balance and attrition tests as pre-specified in Haushofer et al. (2017a), showing that the randomization succeeded in observationally comparable groups at Baseline. We refer to those analyses and pre-specified outcomes for testing balance and attrition for earlier rounds of data collected.

We will update these analyses for the sample reached at Endline 3, as attrition in particular may change. To be specific, we follow Egger et al (2022) for analyzing attrition, and whether it differed by treatment status. We estimate Equation (1) using as an outcome an indicator r_{vhs} for whether household h in village v in sublocation s is observed at endline, and do this separately for eligible and ineligible households, and with r_{vhs} defined either as being reached at Endline 3, in both Baseline and Endline 3, and at Baseline, Endline 1, Endline 2, and Endline 3. If we find worrying levels of differential attrition, we will adjust for potential bias by bounding the parameter of interest using Lee Bounds (Lee 2009) or more recent econometric innovations (e.g. Semenova 2020) and by using a weighted least squares estimator with the inverse probability of selection as weights.

5.6 Heterogeneous impacts

In addition to heterogeneity by timing and cause of death that have been noted above, we will also make use of several dimensions of the data to look at heterogeneity in treatment effects. We note that the household census data collects a much more limited set of variables than the household surveys, so we will not be able to estimate effects on all of the same dimensions that have been previously pre-specified

(e.g. in Haushofer et al. 2017). We can and will explore heterogeneity by (i) child gender, (ii) maternal age (specifically, the bins that we have outlined in the specifications above), and (iii) by child birth order.

We follow the approach of Haushofer et al. (2017) (interacting indicators for these variables with the treatment term) for these dimensions of heterogeneity. We may also run analyses for other sources of heterogeneity in a more exploratory manner.

For this analysis, the primary focus will be on effects for recipient households, but as a secondary focus we will also investigate heterogeneous effects for non-recipient households.

6 References

Abdul Latif Jameel Poverty Action Lab (J-PAL). 2020. "Using cash transfers to improve child health in low- and middle-income countries." J-PAL Policy Insights. Last modified May 2020. https://doi.org/10.31485/pi.2523.2020

Amek, N.O., Van Eijk, A., Lindblade, K.A. et al. Infant and child mortality in relation to malaria transmission in KEMRI/CDC HDSS, Western Kenya: validation of verbal autopsy. Malar J 17, 37 (2018). https://doi.org/10.1186/s12936-018-2184-x

Anderson, M.L., 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecederian, Perry Preschool and Early Training Projects", Journal of the american Statistical Association, December 2008, Vol. 103, No. 484: 1481-1495.

Byass, P., Herbst, K., Fottrell, E., Ali, M. M., Odhiambo, F., Amek, N., Hamel, M. J., Laserson, K. F., Kahn, K., Kabudula, C., Mee, P., Bird, J., Jakob, R., Sankoh, O., & Tollman, S. M. (2015). Comparing verbal autopsy cause of death findings as determined by physician coding and probabilistic modelling: a public health analysis of 54 000 deaths in Africa and Asia. Journal of global health, 5(1), 010402. https://doi.org/10.7189/jogh.05.010402

Benjamini, Y., A. M. Krieger, D. Yekutieli, 2006. "Adaptive Linear Step-up Procedures That Control the False Discovery Rate", Biometrika, Volume 93, Issue 3, September 2006, Pages 491–507, https://doi.org/10.1093/biomet/93.3.491

Chernozhukov, V., M. Demirer, E. Duflo and I. Fernández-Val, 2018. "Generic Machine Learning Inference on Heterogeneous Treatment Effects in Randomized Experiments, with an Application to Immunization in India", NBER working paper.

Egger, D., J. Haushofer, E. Miguel, P. Niehaus and M. Walker, 2022. "General equilibrium effects of cash transfers: Experimental evidence from Kenya", *Econometrica* 90(6): 2603-2643. https://doi.org/10.3982/ECTA17945

Egger, D., J. Haushofer, E. Miguel, and M. Walker, 2021a. "GE Effects of Cash Transfers: Pre-analysis plan for Endline 2 Household", July 2021. https://www.socialscienceregistry.org/trials/505

Egger, D., J. Haushofer, E. Miguel, and M. Walker, 2021b. “GE Effects of Cash Transfers: Pre-analysis plan for Endline 2 Local Economy Analyses”, December 2021. AEA Trial Registry: https://www.socialscienceregistry.org/trials/505

Gacheri, S., Kipruto, H., Amukoye, E., Ong, J., Mitchell, E. M. H., Sitienei, J., Kiplimo, R., & Muturi, C. (2014). Performance of clinicians in identifying tuberculosis as cause of death using verbal autopsy questionnaires in Siaya County, Kenya. African Journal of Health Sciences, 27, 232-238.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, *225*(2), 254–277. https://doi.org/https://doi.org/10.1016/j.jeconom.2021.03.014

Haushofer, J., E. Miguel, P. Niehaus and M. Walker. 2017a. “General Equilibrium Effects of Cash Transfers: Pre-analysis Plan for household welfare analysis”, July 2017. AEA Trial Registry: https://www.socialscienceregistry.org/trials/505

Haushofer, J., E. Miguel, P. Niehaus and M. Walker. 2017b. “General Equilibrium Effects of Cash Transfers: Pre-analysis Plan for Targeting Analysis”, July 2017. AEA Trial Registry: https://www.socialscienceregistry.org/trials/505

Lee, D. S. 2009.“Training, Wages and Sample Selection: Estimating Sharp Bounds on Treatment Effects”, The Review of Economic Studies 76 (3): pp. 1071-1102.

Murray, C.J., Lopez, A.D., Black, R. et al. Population Health Metrics Research Consortium gold standard verbal autopsy validation study: design, implementation, and development of analysis datasets. Popul Health Metrics 9, 27 (2011). https://doi.org/10.1186/1478-7954-9-27

Nyaguara O. Amek, Frank O. Odhiambo, Sammy Khagayi, Hellen Moige, Gordon Orwa, Mary J. Hamel, Annemieke Van Eijk, John Vulule, Laurence Slutsker & Kayla F. Laserson (2014) Childhood cause-specific mortality in rural Western Kenya: application of the InterVA-4 model, Global Health Action, 7:1, DOI: 10.3402/gha.v7.25581

Odhiambo FO, Laserson KF, Sewe M, Hamel MJ, Feikin DR, Adazu K, Ogwang S, Obor D, Amek N, Bayoh N, Ombok M, Lindblade K, Desai M, ter Kuile F, Phillips-Howard P, van Eijk AM, Rosen D, Hightower A, Ofware P, Muttai H, Nahlen B, DeCock K, Slutsker L, Breiman RF, Vulule JM. Profile: the KEMRI/CDC Health and Demographic Surveillance System--Western Kenya. Int J Epidemiol. 2012 Aug;41(4):977-87. doi: 10.1093/ije/dys108. PMID: 22933646.

Romano, J. P., and M. Wolf. 2005a. Exact and Approximate Stepdown Methods for Multiple Hypothesis Testing. *Journal of the American Statistical Association* 100(469): 94–108.

———. 2005b. Stepwise Multiple Testing as Formalized Data Snooping. *Econometrica* 73(4): 1237–1282.

———. 2016. Efficient computation of adjusted p-values for resampling-based stepdown multiple testing. *Statistics and Probability Letters* 113: 38–40.

Semenova, Vira. *Generalized Lee Bounds* (2020). https://arxiv.org/abs/2008.12720

Serina, P., Riley, I., Stewart, A. et al. Improving performance of the Tariff Method for assigning causes of death to verbal autopsies. BMC Med 13, 291 (2015). https://doi.org/10.1186/s12916-015-0527-9

Walker, M. (2017). "Pre-Analysis Plan: Local Public Finance and Unconditional Cash Transfers in Kenya." February 2017. AEA Trial Registry: https://www.socialscienceregistry.org/trials/505