Distribution Agreement

Signature:

______________________________ ______________

Katherine Stanfill Date

Use of Capture-Recapture to Assess Poor Pregnancy Outcomes in a Tanzanian Refugee Camp

By

Katherine Stanfill

MPH

Global Health

____

Michelle Hynes, PhD, MPH

Committee Chair

____

Erin Hulland, MPH

Committee Member

Contents

Abstract

Poor pregnancy outcomes, including neonatal mortality, stillbirths, and abortions, are exacerbated in populations effected by complex humanitarian emergencies (CHEs).

The Nyarugusu refugee camp located in northwest Tanzania was established in 1996 and is home to approximately 70,000 refugees, 17,500 of whom are women of reproductive age (15-49 years old). Surveillance on pregnancy outcomes is collected through the camp's health information system (HIS). In order to examine how well the camp's HIS is capturing these events, HIS pregnancy outcome data over a 1-year period was compared to pregnancy outcome data from a household survey covering the same period.

A secondary data analysis was conducted on the two data sources utilizing the methodology of record linkage matching and capture-recapture surveillance. For the period of June 20, 2012 to June 20, 2013, household surveys captured 1,883 live births, 5 neonatal deaths, 31 still births and 50 abortions. The HIS captured 1,555 live births, 3 neonatal deaths, 26 stillbirths and 49 abortions. Probabilistic record linkage was utilized to assign similarity scores to identified cases within each database and those above a determined threshold are determined to be matches. Capture-recapture was utilized in order to estimate the number of events in the population. Calculated rates for each of the pregnancy outcomes within the camp's population are as follows: 3.7 neonatal deaths per 1,000 live births, 50.5 stillbirths per 1,000 live births, and 270.3 abortions per 1,000 live births.

Analyzing the HIS in conjunction with a household survey revealed that there were high levels of non-matches between the datasets. The identification of cases on one dataset but not the other indicates that neither dataset is fully capturing the events. Therefore, capture-recapture is useful for surveillance evaluation and strengthening the HIS is recommended. Elements of the household surveys can be incorporated into the standard surveillance, such as utilizing community health workers to capture non-facility events.

Chapter 1: Literature Review

Global burden of poor pregnancy outcomes

Neonatal mortality:

In Sub-Saharan Africa, 1 in 9 children die before age five, a rate that is more than 16 times the average for developed regions. In an updated systematic analysis of global, regional, and national causes of child mortality, researcher Liu states, “If the present trends continue, 4.4 million children younger than 5 years will still die in 2030. Furthermore, sub-Saharan Africa will have 33% of the births and 60% of the deaths in 2030.” (Liu, Oza et al. 2015) Sub-Saharan Africa also has the highest risk of death in the first month of life and is among the regions showing the least progress (Liu, Oza et al. 2015). Neonatal mortality, defined as a death within the first 28 days of a child’s life, accounted for 2.7 million deaths in 2015. Of those, almost 1 million of the neonatal deaths occurred on the day of birth, and close to 2 million died within the first week of life (Filippi, Ronsmans et al. 2006). The proportion of neonatal deaths of all under-five deaths has increased in all World Health Organization (WHO) regions over the last 25 years, which include the African Region, Region of the Americas, South-East Asia Region, European Region, Eastern Mediterranean Region, and Western Pacific Region. The neonatal mortality rates have not decreased as quickly when compared to children aged 1-5. (Filippi, Ronsmans et al. 2006).

Research shows that up to two-thirds of neonatal deaths can be prevented by quality care at time of birth and within the first week of life. (WHO 2016) Interventions to reduce neonatal mortality differ from those designed to reduce under-five mortality, as they are specific to the neonatal time period. Causes of death that are closely linked to the neonatal period pertain to the care of the mother and child around the time of pregnancy (UN. 2015). Evidence-based, cost-effective interventions include quality care during pregnancy, safe and clean delivery by a skilled birth attendant, immediate postnatal care, neonatal resuscitation, extra care of low birth weight babies, attention to baby warmth, treatment of neonatal sepsis, and early initiation of breastfeeding (Wardlaw, You et al. 2014)

Abortions and Stillbirths:

In addition to neonatal mortality, additional poor pregnancy outcomes include stillbirths and abortions. A stillbirth is defined by the WHO as, “a baby born with no signs of life at or after 28 weeks’ gestation.” (WHO 2016) Worldwide, the number of stillbirths has declined by 19.4% between 2000 and 2015, although there remained a total of 2.6 stillbirths globally in 2015. Ninety-eight percent of these occurred in low- and middle-income countries, with a rate of 29 stillbirths per 1000 births in sub-Saharan Africa. (WHO 2016) It is also known that the majority of stillbirths are preventable, with rates directly correlating with access to maternal healthcare. However, recommended interventions are not standardized and data on tracking stillbirths is limited. (Fretts 2005) The Lancet article *Stillbirths: progress and unfinished business* calls for action in that, “a universally adopted definition of stillbirth is needed in addition to a global repository to access such data.” (Frøen, Friberg et al. 2016) Additionally, a miscarriage is the spontaneous removal of a fetus or embryo, resulting in the termination a pregnancy. If it is caused purposely, it is known as an induced abortion. The number of woman having abortions is slightly elevated for those living in a developing region, with the overall rates in Africa showing no decline between 2003 and 2008, with approximately 29 abortions per 1,000 women of child-bearing age. (WHO 2012)

Global poor pregnancy outcome reduction goals and progress:

In 2000, the United Nations (UN) established the Millennium Development Goals (MDGs) that aimed to focus global healthcare efforts throughout the following decade and through the year 2015. This UN initiative established targets for addressing extreme poverty, education, gender equality, child mortality, maternal health, HIV/AIDS, malaria, and environmental sustainability. (UN. 2015). Within the MDGs, Target 4.A states, “reduce by two-thirds, between 1990 and 2015, the under-five mortality rate” (UN. 2015). In 1990 there were 12.7 million under-five deaths globally (91 deaths per 1,000 live births). By 2015, the under-five mortality rate had fallen 53% to 5.9 million deaths (43 deaths per 1,000), an achievement just short of the MDG goal (UN. 2015). Within this same time period, neonatal mortality decreased by 42%, resulting in an increase in the overall percent of neonatal deaths among all under 5 deaths (Taylor, Williams et al. 2015). The MDGs did not track stillbirths specifically and progress made towards reducing stillbirths was slower than for neonatal deaths. (Blencowe, Cousens et al. 2016)

In 2015, as the MDGs came to an end, the UN revised the goals into the new Sustainable Development Goals (SDGs). The SDGs differentiate between under-five mortality rates and neonatal mortality rates. The SDGs state, "By 2030, end preventable deaths of newborns and children under 5 years of age, with all countries aiming to reduce neonatal mortality to at least as low as 12 per 1,000 live births and under-five mortality to at least as low as 25 per 1,000 live births" (Liu, Oza et al. 2015). More specific poor pregnancy outcomes, including abortion and stillbirths are broadly encompassed within the 'preventable deaths of newborns and children under 5 years of age' with no specific targets outlined. Progress made towards the SDGs is underway and is monitored by the UN.

Complex Humanitarian Emergencies and poor pregnancy outcomes:

Poor pregnancy outcomes are exacerbated in populations affected by complex humanitarian emergencies (CHEs). The UN defines a CHE as, "a humanitarian crisis in a country, region, or society where there is a total or considerable breakdown of authority resulting from internal or external conflict and which requires an international response that goes beyond the mandate or capacity of any single and/or ongoing UN country program." (Brennan and Nandy 2001) Natural disasters, civil wars, or political unrest, often result in the increased need of humanitarian assistance in circumstances where delivery is challenging. As a result, there are often, "populations affected by armed conflict have experienced severe public health consequences mediated by population displacement, food scarcity, and the collapse of basic health services" (Toole and Waldman 1997). These populations could be internally displaced persons (IDPs), meaning persons who have been forced to flee their home but have remained within their country's borders, or refugees, meaning persons who are outside their country of citizenship. A refugee could have been forced to leave in order to escape war, persecution, or natural disaster. The United Nations High Commission for Refugees (UNHCR) is an agency mandated to protect and support refugees, including those who have survived complex humanitarian emergencies.

At the end of 2014, there were 3.7 million refugees in sub-Saharan Africa, which was an increase of 759,000 refugees from the previous year. International aid responded covering many different focus areas. In regards to poor pregnancy outcomes, "United Nations agencies provided reproductive health services in emergencies, quickly transitioning into more comprehensive programs in refugee areas." (UN 2015) This was in response to evidence based research indicating the exacerbation of poor pregnancy outcomes within these settings. It has been found that neonatal deaths have been exacerbated in particular. For example, "In 1991, countries with high levels of political instability accounted for approximately 10% of all neonatal deaths worldwide; in 2013, this figure had grown to 31%" (Wise and Darmstadt 2015).

Within refugee camp settings, there is often designated healthcare through humanitarian aid, often under the jurisdiction of UNHCR or other international groups. This healthcare provided in times of emergency for a specified population may be of higher quality that the baseline healthcare within the local population of the geographic area. (Brennan and Nandy 2001) This could be due to a myriad of factors such as an increase of healthcare funding for the particular situation, outside medical expertise coming into the refugee camp, greater accessibility of health services, or the prioritization of providing health services to the displaced population. (Brennan and Nandy 2001) Data of health outcomes often indicate better outcomes within the camp setting as compared to the host or home countries, where conflict is often exacerbated (O'Heir 2004). While obstetric services in general are present in these healthcare systems, neonatal care is not frequently integrated into obstetric care. In addition, pregnant women and their families do not necessarily have the skills to recognize life-threatening conditions and/or are able to seek health care in time (O'Heir 2004). Largely, these deficiencies demonstrate that despite an additional emphasis on health care within refugee camps, care specifically for neonates is generally still a low priority.

HIS within Refugee camps:

Within refugee camps, health care services and health outcomes are tracked. In order to do this, health data is systematically collected, through a health information system (HIS). HIS consists of the cohesive collection of health care services data, health outcomes data, management of this information, analysis for dissemination to all stakeholders, and decision making in camp planning. This information and data is then utilized for informing programming and allocation of resources within the camp. Ideally, camp programs should base their health interventions on objective epidemiological data (Brennan and Nandy 2001).

Specifically within post-emergency refugee camp settings, HIS are under the jurisdiction of UNHCR (Nobuko, Lale et al. 2012). In order to track progress towards the SDGs, "UNHCR and its partners have developed and implemented a standardized health information system in numerous countries to monitor camp-based refugee health programs" (Haskew, Spiegel et al. 2010). Standardizing this process enables ease of data analysis comparable across the globe, thus accurate reporting of progress towards the SDGs. Within the refugee camps, the HIS presents data indicating a need for increased resources, logistics, staffing, and funding. (Haskew, Spiegel et al. 2010). This data therefore provides evidence for shaping programming and allocating resources.

Challenges of poor pregnancy outcome surveillance in HIS:

Challenges remain in capturing poor pregnancy outcome data within health information systems in post-emergency refugee camp settings. Births frequently occur outside of the healthcare/clinic locations where there is no data collection, leaving uncaptured health events out of the surveillance system. Lack of resources presents another challenge. Comprehensive, well-functioning surveillance systems require finances, personnel, and upkeep that are often unavailable in a refugee camp strapped for funding. Fluctuations in the population count contribute to the challenges. Growing populations put strains on even basic resources such as food or water within a refugee camp setting, creating an environment where it is difficult to provide sustained care over a period of time (Hynes, Sakani et al. 2012).

Overall, monitoring health outcomes remains weak, which leaves a need for the use of statistical models to obtain estimates for most countries (Nobuko, Lale et al. 2012). With the rapidly evolving environment of CHEs, the camp population is dynamic. Therefore, the data capture needs to be scalable in order to allow for transition within the population, a variable that is difficult for a static health information system to take into consideration. The UNHCR HIS has been hampered by a lack of understanding of differences in definitions, statistical tabulations, and reporting requirements among clinical providers, despite their standardized reference manual, demonstrating caveats between intention and implementation (Barfield 2011). All existing data sources have limitations, and therefore utilizing an analysis using multiple sources has the potential to better estimate neonatal deaths. Historically, the gold-standard for understanding poor pregnancy outcomes captured by the HIS is through capture-recapture methodology. Capture-recapture is utilized in ascertaining health outcomes in post-emergency refugee camp settings. Its purpose is to estimate the number of events in a population which is done by utilizing multiple data sources, the first of which is referred to as the "capture" and the following of which are referred to as the "recapture". In its most simplistic form, capture-recaptures is conducted using two data sources. (Hook and Regal 1995)

Refugee Situation in Tanzania:

The Nyarugusu refugee camp, located in northwest Tanzania, was established in 1996 and consists of predominately Burundian refugees and those coming from Democratic Republic of Congo's South Kivu Province. The camp encompasses an area of 26.56 square miles and is home to approximately 70,000 refugees, 17,500 of whom are women of reproductive age (15-49 years old). The refugee camp is organized by the UNHCR and services are provided in collaboration with other partner organizations. UNHCR's primary purpose is to safeguard the rights and well-being of refugees. UNHCR published a report detailing that in 2012, the health situation in the camp remained stable, with the under-five mortality rates at average levels (UNHCR 2013). The camp is served by two hospitals, referred to as the Dispensary and the Annex. The Dispensary is the main hospital complex, located at the center of the camp and includes a labor and delivery ward, operating theater, inpatient adult female ward, inpatient adult male ward, and a pediatric ward. The Annex is the smaller hospital complex within the camp, located north of the Dispensary, and consists of a labor and delivery ward, an adult female inpatient ward, inpatient male adult ward, and a pediatric ward. These hospitals are operated and staffed by the Tanzanian Red Cross Society. In addition, the camp also employs traditional healers from the refugee community who serve as community health workers (CHWs) who are responsible for addressing routine health needs of refugees living in their coverage area. Within the camp, home births are discouraged and CHWs are trained to refer women in need of services to the hospitals. The HIS within the camp is established by UNHCR, run by the Tanzanian Red Cross, and is influenced by the challenges of operating within such a population. Few details are known on the neonatal mortality statistics within the population and analyses must take into context the known information of similar settings and host country and country-of-origin population statistics. (UNHCR 2013)

Chapter 2: Embedded Manuscript

Abstract

Poor pregnancy outcomes, including neonatal mortality, stillbirths, and abortions, are exacerbated in populations effected by complex humanitarian emergencies (CHEs).

The Nyarugusu refugee camp located in northwest Tanzania was established in 1996 and is home to approximately 70,000 refugees, 17,500 of whom are women of reproductive age (15-49 years old). Surveillance on pregnancy outcomes is collected through the camp's health information system (HIS). In order to examine how well the camp's HIS is capturing these events, HIS pregnancy outcome data over a 1-year period was compared to pregnancy outcome data from a household survey covering the same period.

A secondary data analysis was conducted on the two data sources utilizing the methodology of record linkage matching and capture-recapture surveillance. For the period of June 20, 2012 to June 20, 2013, household surveys captured 1,883 live births, 5 neonatal deaths, 31 still births and 50 abortions. The HIS captured 1,555 live births, 3 neonatal deaths, 26 stillbirths and 49 abortions. Probabilistic record linkage was utilized to assign similarity scores to identified cases within each database and those above a determined threshold are determined to be matches. Capture-recapture was utilized in order to estimate the number of events in the population. Calculated rates for each of the pregnancy outcomes within the camp's population are as follows: 3.7 neonatal deaths per 1,000 live births, 50.5 stillbirths per 1,000 live births, and 270.3 abortions per 1,000 live births.

Analyzing the HIS in conjunction with a household survey revealed that there were high levels of non-matches between the datasets. The identification of cases on one dataset but not the other indicates that neither dataset is fully capturing the events. Therefore, capture-recapture is useful for surveillance evaluation and strengthening the HIS is recommended. Elements of the

household surveys can be incorporated into the standard surveillance, such as utilizing community health workers to capture non-facility events.

Background

In Sub-Saharan Africa, 1 in 9 children die before age five, a rate that is more than 16 times the average for developed regions. Neonatal mortality, the death within the first 28 days of a child's life, accounted for 2.7 million deaths in 2015. Up to two-thirds of neonatal deaths can be prevented by quality care at time of birth and within the first week of life. (WHO 2016) In addition to neonatal mortality, additional poor pregnancy outcomes include stillbirths and abortions. A stillbirth is defined by the WHO as, "a baby born with no signs of life at or after 28 weeks' gestation." (WHO 2016) It is also known that the majority of stillbirths are preventable, with rates directly correlating with access to maternal healthcare. However, recommended interventions are not standardized and data on tracking stillbirths is limited. (Fretts 2005) An additional poor pregnancy outcome is when pregnancy ends by removal of a fetus or embryo before surviving the uterus. If this occurs spontaneously, it is known as a miscarriage. If it is caused purposely, it is also known as an induced abortion. A woman's likelihood of having an abortion is slightly elevated if she lives in a developing region, with the overall rates in Africa showing no decline between 2003 and 2008, with approximately 29 abortions per 1,000 women of child-bearing age. (WHO 2012)

Poor pregnancy outcomes are exacerbated in populations effected by complex humanitarian emergencies (CHEs). The United Nations (UN) defines a CHE as, "a humanitarian crisis in a country, region, or society where there is a total or considerable breakdown of authority resulting from internal or external conflict and which requires an international response that goes beyond the mandate or capacity of any single and/or ongoing UN country program." Often resulting from natural disasters, civil wars, or political unrest, humanitarian assistance is much needed and delivery is challenging. The United Nations High Commission for Refugees (UNHCR) is an agency mandated to protect and support refugees such as those who have survived complex humanitarian emergencies. Relevant to poor pregnancy outcomes, "United

Nations agencies provided reproductive health services in emergencies, quickly transitioning into more comprehensive programs in refugee areas." (UN 2015) This was in response to evidence based research indicating the exacerbation of poor pregnancy outcomes within these settings. "In 1991, countries with high levels of political instability accounted for approximately 10% of all neonatal deaths worldwide; in 2013, this figure had grown to 31%" (Wise and Darmstadt 2015).

Within refugee camps, health care services and outcomes are tracked through a health information system (HIS), which is the cohesive collection of data, managing information, analyzing for dissemination, and decision making. Health information systems within post-emergency refugee camp settings are under the jurisdiction of the United Nation's High Commissioner for Refugees (UNHCR). (Nobuko, Lale et al. 2012) Comprehensive, well-functioning surveillance systems require finances, personnel, and upkeep that are often unavailable in a refugee camp strapped for funding. Fluctuations in the population count contribute to the challenges. Overall, monitoring health outcomes remains weak which leaves a need for the use of statistical models to obtain estimates for most countries (Nobuko, Lale et al. 2012). Within the refugee camps, the HIS presents a necessity for increased resources, logistics, staffing, and funding, simply for establishment (Haskew, Spiegel et al. 2010). Historically, the gold-standard for understanding neonatal mortality captured by the HIS and other sources within refugee camps is through capture-recapture methodology.

The Nyarugusu refugee camp located in northwest Tanzania was established in 1996 and consists of predominately Burundian refugees and those coming from Democratic Republic of Congo's South Kivu Province. The camp encompasses an area of 26.56 square miles and is home to approximately 70,000 refugees, 17,500 of whom are women of reproductive age (15-49 years old). UNHCR published a report detailing that in 2012, the health situation in the camp remained stable, with the under-five mortality rates at average levels. (UNHCR 2013) The camp is served by two hospitals, referred to as the Dispensary and the Annex. The Dispensary is the main hospital complex, located at the center of the camp and includes a labor and delivery ward, operating theater, inpatient adult female ward, inpatient adult male ward, and a pediatric ward. The Annex is the smaller hospital complex within the camp, located north of the Dispensary, and consists of a labor and delivery ward, an adult female inpatient ward, inpatient male adult ward, and a pediatric ward. These hospitals are operated and staffed by the Tanzanian Red Cross Society. In addition, the camp also employs traditional healers from the refugee community who serve as community health workers (CHWs) who are responsible for addressing routine health needs of refugees living in their coverage area. Within the camp, home births are discouraged and CHWs are trained to refer women in need of services to the hospitals. Therefore, the HIS within the camp is established by UNHCR, run by the Tanzanian Red Cross, and is influenced by the challenges of operating within such a population. Few details are known on the neonatal mortality statistics within the population and analyses must take into context the known information of similar settings and host country and country-of-origin population statistics. (UNHCR 2013)

In the Nyarugusu refugee camp, birth outcome data is collected via the camp's UNHCR Health Information System (HIS). A twelve month retrospective study (June 20, 2012 – June 20, 2013) was conducted to measure neonatal deaths, stillbirths, abortions and live births. For this study, a household survey was conducted, collecting the camp's birth outcome data. This study involved looking at two sources, HIS and household survey. The HIS included line-by-line review of all HIS registers, log books, and community health worker registers to create a list of all birth-related events inside the camp during the study time frame. The HIS did not collect data throughout the entirety of the 12 month study period. The records have written documentation that no abortion log was kept at the main dispensary post Jan. 2013, leaving no data captured during a portion of the stated study period. In addition, no abortion log was kept whatsoever at the annex. Thus, it is unclear if there were any abortions at the annex. The household survey consisted of face-to-face household interviews that were conducted with every women of reproductive age (15-49 years) inside Nyarugusu refugee camp, with questions on her pregnancy history (neonatal deaths, stillbirths, abortions, live births) in the previous 12 month period. This data was collected by the community health workers (CHWs), traditional birth attendants (TBAs) and the midwives who were trained to conduct the household surveys.

A secondary data analysis was conducted in order to further examine this data, utilizing the methodologies of record linkage matching and capture-recapture. Record linkage is the process of deciding which records from one or more databases refer to the same entity, even when those records do not match exactly on any combination of other identifiers. Probabilistic record linkage is a family of record linkage techniques that assigns similarity scores to pairs of records and treats all pairs that score above a certain threshold as matches. (Glenn Wright 2010) Similarity scores for each pair of records were calculated and those above a determined threshold are determined to be matches.

Capture-recapture is utilized in ascertaining health outcomes in post-emergency refugee camp settings and has the purpose of estimating the number of events in a population. This is done by utilizing two or more data sources, one of which is referred to as the “capture” and the other of which is referred to as the “recapture”. Capture-recapture methodology relies on the following assumptions: independence between data sources, homogeneity of capture probabilities, error-free matching across lists, stable population, and no duplicates within lists. (Hook and Regal 1995) Outcomes identified separately on each list in conjunction with the degree of overlap between the lists are utilized to estimated missed outcomes and overall outcomes within the population. Accuracy of statistical estimates based off of capture-recapture methodology depends on extent of record overlap, the ability for each surveillance system to follow the assumption that an event can equally be captured by each (i.e. the event must be “capturable” by the system), and the event having high frequency (as opposed to being rare) (Hook and Regal 1995).

These methodologies were utilized in order to better understand the capture of poor pregnancy outcomes within the refugee camp HIS, answering the following question: how do the pregnancy outcomes of neonatal mortality, stillbirths, and abortions captured via a one-time household survey compare to those captured via the standard health information system within the Nyarugusu refugee camp in Tanzania over a 12-month recall period?

Methods

Matching

For this secondary data analysis, data utilized was from an exhaustive 12-month retrospective study (June 20, 2012 – June 20, 2013) from two data sources: the camp's Health Information System (HIS) and a household survey.

Health outcomes of interest include neonatal deaths, stillbirths, abortions and live births. Each was defined according to World Health Organization's standard as follows:

Neonatal death: Death in the first 28 completed days of life
Stillbirth: Fetal death after 28 weeks of gestation
Abortion: Pregnancy termination prior to 20 weeks' gestation (both spontaneous and induced)
Live Birth: Complete expulsion or extraction of a product of conception which breathes or shows any sign of life

The raw data of the household survey responses and collected HIS data pertaining to the health outcomes of interest were imported via Excel files into SAS software version 9.3 for data cleaning and analysis. (Institute. 2016)

Linkage

For the secondary analysis of the HIS and household survey data, matching was conducted via probabilistic record linkage as adopted from guidance by Glenn Wright. (Glenn Wright 2010). Each poor pregnancy outcome (neonatal death, abortion, and stillbirth) had its own unique matching conducted and therefore each had its own unique analysis. A list of all included variables for each dataset was examined in order to identify matching fields across both lists. Every variable that was present on both lists was then utilized in the matching process. For this analysis, those variables were date, address, age, and sex. Points were assigned to each variable, with the most specific matching details scoring higher value points and less specific matching details scoring lower value points. Thus, if a matching variable would result in a higher likelihood of indicating a unique match, it received higher points. If a matching variable would result in a lower likelihood of indicating a unique match, such as more common characteristics that many cases could identify with, it received lower points. Table 1 below defines the scoring breakdown of all matching variables:

Table 1: Matched Variables assigned linkage points

Matching Variable	**Congruency**	**Assigned Points**
Date	Missing	0
	Exact Match	20
	Same year, transposed day/month	12
	2 of 3 match (day/month/year)	12
	Else	-2
Address	Missing	0
	Exact match	15
	Else	-2
Age	Missing	0
	Exact match	12
	± one year	9
	Else	-2
Sex	Missing	0
	Exact match	3
	Else	0

The variable of date was individually analyzed for matching and the specificity of matching dates was determined to be of high value. Similarity score points were assigned, full points (20) for an exact match, near-full (12) points for the same year, but transposed day and month, reduced points (12) when only two of the three fields (day, month, year) matched, no (0) points for any cases missing data for the variable, and negative (-2) points for a non-match. The distribution of the similarity scores were sorted and depicted in a histogram. The distribution for each poor pregnancy outcome (neonatal death, abortion, and stillbirth) had its own unique distribution of similarity scores and therefore each had its own unique threshold.

The variable of address was individually analyzed for matching and the specificity of matching dates was determined to be of high value. Similarity score points were assigned, full points (15) for an exact match, no (0) points for any cases missing data for the variable, and negative (-2) points for a non-match. The distribution of the similarity scores were sorted and depicted in a histogram.

The variable of age was individually analyzed for matching and the specificity of matching dates was determined to be of medium value. Similarity score points were assigned, full points (12) for an exact match, partial points (9) for an inexact match of $\pm$ 1 year, no (0) points for any cases missing data for the variable, and negative (-2) points for a non-match. The distribution of the similarity scores were sorted and depicted in a histogram.

The variable of sex was individually analyzed for matching and the specificity of matching dates was determined to be of low value. Similarity score points were assigned, full points (3) for an exact match, no (0) points for any cases missing data for the variable, and negative (-2) points for a non-match. The distribution of the similarity scores were sorted and depicted in a histogram.

Variables were analyzed in a combined fashion to determine a matched total score, including all of the variables identified as matched on each list. Assigned points for each variable individually was re-utilized in the combined analysis, with the similarity scores for the total matched scores equating to the summation of each individual variable points. The distribution of the similarity scores were sorted and depicted in a histogram. Thresholds were determined for each of the poor pregnancy outcomes (neonatal death, stillbirth, and abortion) from the distribution of total similarity scores, based on all matched variables.

In order to capture as many true matches and exclude as many false matches as possible, thresholds were set conservatively to be at similarity score point values that were seen as the highest amongst all of the potential matches. Once this high threshold was established for each outcome, the small subset of observations above the threshold was manually analyzed to check for the presence of true matches.

Capture - Recapture

Each case above the determined threshold for the total similarity score was then individually manually analyzed for matching and deemed a "true match". This process created a list of matches, or a case that was determined to be captured by both datasets and was labeled "observed/observed". Non-matches, or those cases of the poor pregnancy outcomes that were identified within an individual dataset but not within both, were labeled as observed/not observed. The counts of each of these were then entered into a two-by-two table for pairwise matching.

The two-by-two table for pairwise matching then utilized the Chapman Estimator to determine an estimation of the count of cases that were not captured by either dataset. The Chapman Estimator, an estimator of population size for capture-recapture studies with two sources, was chosen as it is known to be less biased than other estimators, less affected by zeros, and it is also specifically designed for two sources. (Brittain and Böhning 2009) From the Chapman Estimator, the predicted count of cases that were not captured by either dataset was calculated, along with a predicted total count present for the population as a whole with standard deviation and 95% confidence intervals.

Results

Throughout June 20, 2012 – June 20, 2013, the household surveys captured 1,883 live births and the HIS captured 1,555 live births. There were 5 neonatal deaths identified via household surveys and 3 in the HIS. The household surveys captured 31 stillbirths and 50 abortions whereas the HIS captured 26 stillbirths and 49 abortions.

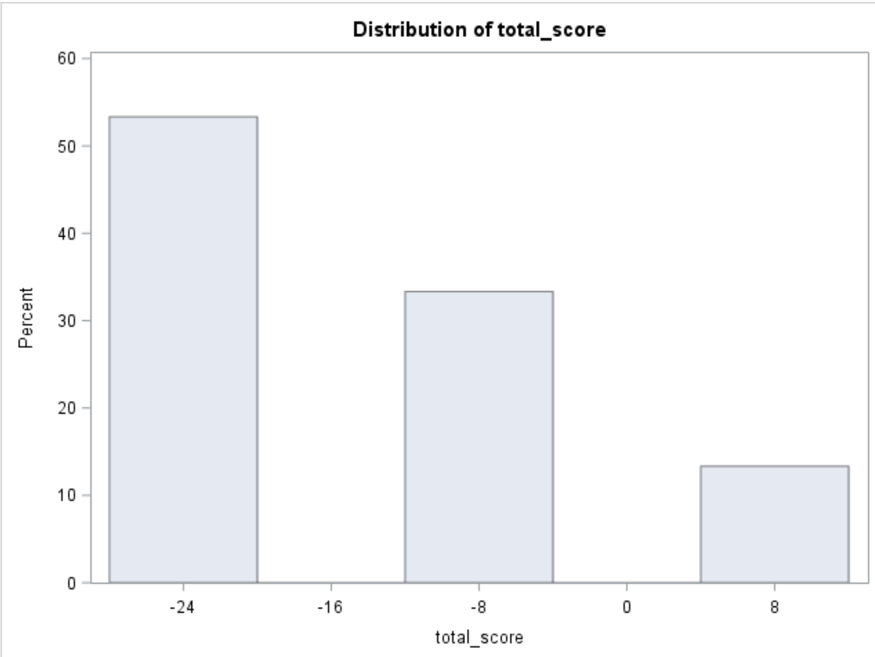
Table 2: Outcomes of Interest by Database

	Household Surveys	HIS
Neonatal Death	5	3
Stillbirth	31	26
Abortion	50	49
Livebirth	1,883	1,555

The HIS had 50 variables with 2,044 observations. The household surveys had 36 variables with 14,022 observations.

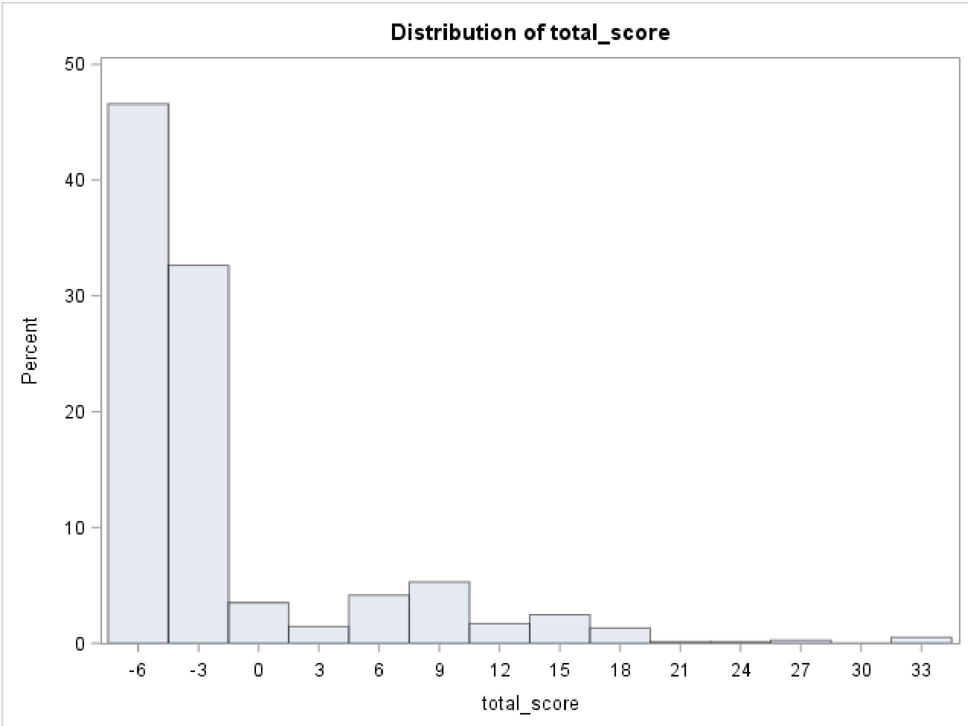
The distribution of total similarity scores for neonatal mortality cases that were above zero were all equal to a value of 8 points. The count of cases resulting in a similarity score of 8 points is 2. Therefore, the threshold, conservatively set at the highest value of identified matches, was set at a total_score value of 8, with a count of 2 cases above threshold.

Table 3: Neonatal Mortality Total Similarity Score Histogram



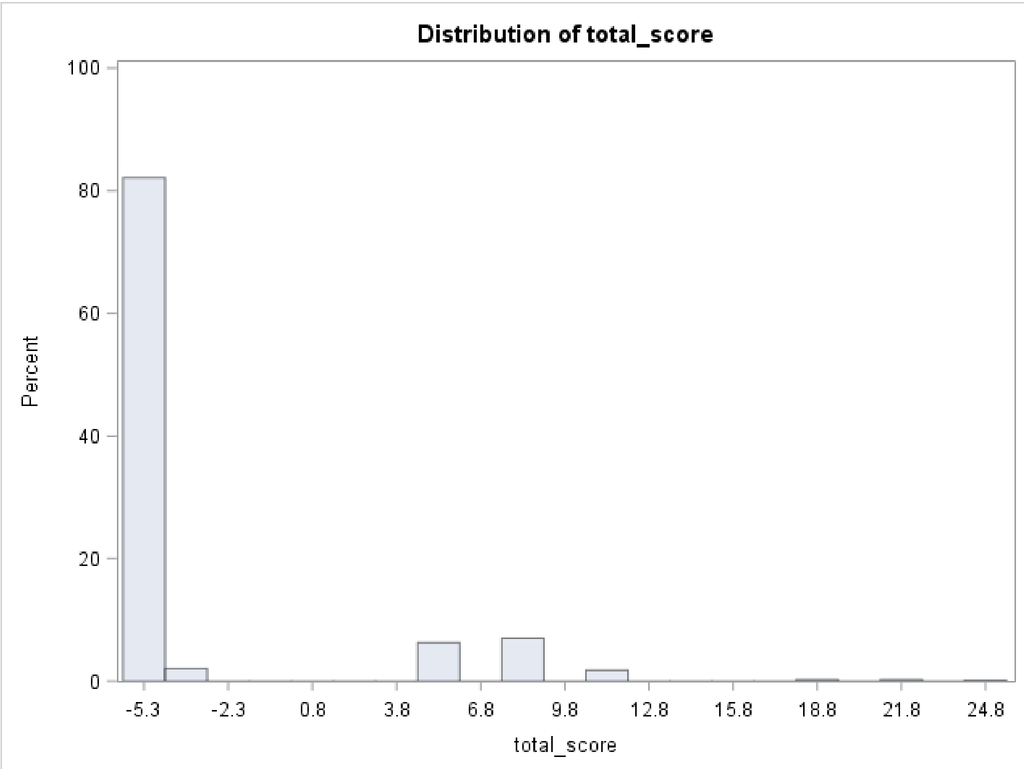
The distribution of total similarity scores for stillbirth cases that were above zero follow a right-skewed distribution. The highest value of the total similarity score identified was 33. Therefore, the threshold, conservatively set at the highest value of identified matches, was set at a total_score value of 33, with a count of 8 cases above threshold.

Table 4: Stillbirth Total Similarity Score Histogram



The distribution of total similarity scores for abortion cases that were above zero follow a right-skewed distribution. The highest value of the total similarity score identified was 24. Therefore, the threshold, conservatively set at the highest value of identified matches, was set at a total_score value of 24, with a count of 4 cases above threshold.

Table 5: Abortion Total Similarity Score Histogram



There were 2 cases of neonatal mortality observed in both the household surveys and the HIS. The household survey and HIS each captured a total of 5 neonatal mortality cases; for both sources, 2 cases were overlaps while 3 cases were unique to each source With a total of 5 cases of neonatal mortality specifically observed within the household surveys, 3 of those cases were not observed within the HIS. With a total of 3 cases of neonatal mortality specifically observed within the HIS, 1 of those cases were not observed within the household surveys. Utilizing the Chapman Estimator, there are an estimated total of 7 (95% Confidence Interval: 4.23- 9.77) cases of neonatal mortality within the population at large. The number of cases not observed in either dataset (d) was calculated utilizing the total count (N) calculated from the Chapman Estimator and subtracting the number of cases observed throughout the datasets. Therefore, a calculated 1 case was not observed within either database.

Table 6: Neonatal Mortality 2x2 Table of Pairwise Matching

	Household Surveys			
HIS		**Observed**	**Not Observed**	
	Observed	2	1	3
	Not Observed	3	d	
		5		

N = 7
SD = 1.41
(4.23, 9.77)

d = 7-2-3-1 = 1

There were 8 cases of stillbirth observed in both the household surveys and the HIS. With a total of 31 cases of stillbirth specifically observed within the household surveys, 23 of those cases were not observed within the HIS. With a total of 26 cases of stillbirth specifically observed within the HIS, 18 of those cases were not observed within the household surveys. Utilizing the Chapman Estimator, there are an estimated total of 95 (95% CI: 53.81 – 136.19) cases of stillbirth within the population at large. A calculated 46 cases were not observed within either database.

Table 7: Stillbirth 2x2 Table of Pairwise Matching

	Household Surveys			
HIS		**Observed**	**Not Observed**	
	Observed	8	18	26
	Not Observed	23	d	
		31		

N = 95
SD = 21.01
(53.81, 136.19)

d = 95-8-23-18 = 46

There were 4 cases of abortion observed in both the household surveys and the HIS. With a total of 50 cases of abortion specifically observed within the household surveys, 46 of those cases were not observed within the HIS. With a total of 49 cases of abortion specifically observed within the HIS, 45 of those cases were not observed within the household surveys. Utilizing the Chapman Estimator, there are an estimated total of 509 (95% CI: 141.32 – 876.68) cases of abortion within the population at large. A calculated 414 cases were not observed within either database.

Table 8: Abortion 2x2 Table of Pairwise Matching

		Household Surveys		
HIS		**Observed**	**Not Observed**	
	Observed	4	45	49
	Not Observed	46	414	
		50	d	

N = 509
SD = 187.59
(141.32, 876.68)

d = 509-4-46-45 = 414

Discussion

As the household survey was designed to cover every woman of reproductive age, a higher number of captured events was expected in comparison to the HIS, especially with HIS data not being collected throughout the entirety of the 12 month study period.

Matching analysis indicated that there were cases identified by both household surveys and through the HIS. However, there were cases identified by household surveys that were not captured by the HIS, and vice versa. In fact, for each poor pregnancy outcome, the counts of overlapping cases were consistently lower than the counts of cases identified within each dataset individually. This indicates there were low levels of matching and that there were few cases where it was confirmed to be the same case that was present within the second dataset. This also indicates that there were high levels of non-matches and that there were individually unique cases present within each dataset. The identification of cases on one dataset but not the other indicates that neither dataset is fully capturing the events. Therefore, the capture-recapture analysis identified poor pregnancy outcomes that no single dataset could do alone.

The total case counts estimated via the Chapman estimator using capture-recapture methodology within the population as a whole were 7 neonatal deaths, 95 stillbirths, and 509 abortions. With a camp population of 70,000 refugees during the study period and approximately 17,500 women of reproductive age (15-49 years old), as well as an identified 1,883 live births captured by household surveys and an identified 1,555 live births captured by the HIS, we calculated the rate of poor pregnancy outcomes as follows: the neonatal death rate was 3.7 deaths / 1,000 live births, the stillbirth rate was 50.5 stillbirths / 1,000 live births, and the abortion rate was 270.3 abortions / 1,000 live births. We conservatively used the figure obtained from the household survey as our denominator as they captured a larger number of live births than HIS, which falls in line with logic as the surveys were comprehensive in nature, interviewing all women of reproductive age, whereas the HIS only covered those seeking out services from the health centers.

In comparison, similar statistical figures have been identified for a similar post-emergency refugee camp, in Chad, where it was estimated for the year of 2011-2012 that there were 46.8 neonatal deaths per 1,000 live births and 44.6 stillbirths per 1,000 live births. (Idowu 2014) Also, WHO has published a report on the United Republic of Tanzania that the neonatal rate for the year 2013 was 21 per 1,000 live births and for the year 2009, the stillbirth rate was 26 per 1,000 live births. (WHO 2013)

Therefore, the rate of neonatal deaths calculated in this analysis is much lower than comparable rates. In fact, the calculated confidence interval for neonatal mortality does not encompass the rates seen in the Chad refugee camp or the country of Tanzania as a whole. This indicates underreporting and that there is likelihood that the rate of neonatal mortality within the Nyarugusu camp is actually higher than what was calculated within this analysis. This could be due to a number of women not seeking out medical facilities when in need of care. Additionally, this could indicate that the threshold set for the matching was set too conservatively high and thus there is a likelihood of more matches if the threshold is set lower. However, specific to the neonatal mortality, neither dataset individually captured enough neonatal deaths to be within range of the rates seen in other similar settings, so even if every case matched, the count would still be low. Therefore, this indicates that neither the household surveys nor the HIS properly captured the poor pregnancy outcome of neonatal death. The number of neonatal deaths likely missed in this analysis is likely to be much higher than what was estimated with the Chapman Estimator. Alternatively, this calculated lower rate of neonatal mortality could be due to improved health services provided by the refugee camp. In a research study on reproductive health indicators and outcomes in these settings, Hynes found that there were lower rates of neonatal mortality within the post-emergency refugee camp as compared to the refugees' country of origin and the host country (Hynes, Sheik et al. 2002).

The rate of stillbirths calculated for the Nyarugusu camp is comparable to other rates, with the calculated confidence interval encompassing the stillbirth rates of the Chad refugee camp and the country of Tanzania as a whole. In fact, the calculated 95% confidence interval indicates that the true rate of stillbirths has a likelihood of being higher than the similar rates.

Therefore, the capture-recapture sheds light on the likelihood of a higher presence of stillbirth than what has been traditionally calculated in similar locations.

There is no comparable abortion rate calculated from the Chad refugee camp or the country of Tanzania as a whole. However, the high rate of abortions could be due to the inclusion of both spontaneous and induced abortions. In addition, there may be other abortions occurring outside of the facilities which are not being captured.

There are a number of limitations to this analysis. First, a limitation is that the health outcomes being analyzed are rare within the population. Rare events within the capture-recapture methodology make the analysis less accurate as the frequency is not high. Another limitation is that the two data lists were not necessarily equally "capturing" the same health outcomes. The household surveys comprehensively and proactively attempted to capture all health outcomes throughout the entire population of interest by interviewing all women of reproductive age. The HIS data list consisted of the health outcomes that came into the HIS system, leaving out those that did not visit a clinic, etc. The HIS also did not collect data throughout the entirety of the 12 month study period. The records have written documentation that no abortion log was kept at the main dispensary post Jan. 2013, leaving no data captured during a portion of the stated study period. In addition, no abortion log was kept whatsoever at the annex. Thus, it is unclear if there were any abortions at the annex. A limitation within the household surveys is recall bias. The census involved interviewing women of reproductive age, inquiring about health information that required a recall period of 12 months. There is also the limitation of misclassification bias. The classification between miscarriage and stillbirth and the classification between stillbirth and neonatal death were explicitly outlined within the analysis, but not at the time of recording said events within the data sources. Additionally, fluctuations of the refugee camps population greatly affect the outcomes of this analysis. An assumption of the capture-recapture utilized is that the population remains stable. While this assumption was made for analysis purposes, there is no feasibility in having a controlled population with no fluctuations within a live refugee camp setting.

In addition, record linkage was a challenge as there were few variables identified in both datasets that were matched. This resulted in many instances of similarity scores indicating false matches. Therefore, conservative high thresholds were set and matches above the threshold were manually analyzed for true/false matches. Thresholds set in this analysis greatly affected the results. If the thresholds had been even slightly less conservatively, the counts of matches identified would have been much higher, resulting in much higher estimated population totals. Therefore, results are sensitive to the thresholds set and this should be kept in mind as there is potential for great variability dependent upon this step. In order to identify more true matches, additional matching should be completed on more specific, unique variables. Specifically, matching on names, a highly specific variable, would provide greater numbers of unique matches. A name is far more likely to be unique to a case and therefore adding specificity, as compared to other characteristics for which a variable could be present for a greater number of cases.

Conclusion

In conclusion, each dataset (HIS and household survey) alone is flawed and the use of two datasets helps to better capture poor pregnancy outcomes. Births and other pregnancy outcomes frequently occur outside of the healthcare/clinic locations where there is no data collection, leaving uncaptured health events out of the surveillance system. Additionally, a potential lack of resources could be a reason as to why the HIS is not fully capturing these events. Comprehensive, well-functioning surveillance systems require finances, personnel, and upkeep. The use of community health workers outside the facilities may not be in place or may be under-utilized due to lack of workers, or finances to employ for these services.

Therefore, capture-recapture is useful for surveillance evaluation and strengthening the HIS is recommended. Elements of the household surveys can be incorporated into the standard surveillance, such as utilizing community health workers to capture non-facility events. Specifically for pregnancy related events, community health workers can be utilized to strengthen the HIS by incorporating home visits as a follow up to women giving birth within the health facilities.

Appendix

Additional Tables

Table 9: Neonatal Mortality Date Similarity Score Histogram

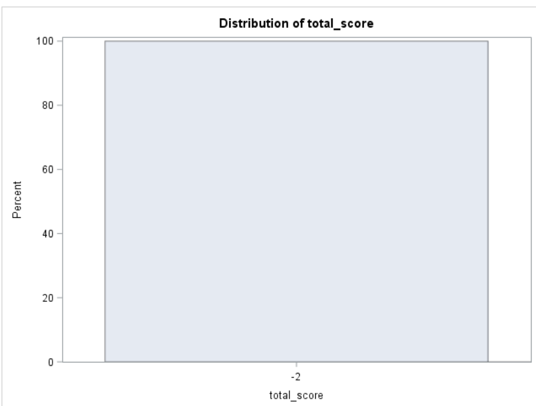


Table 10: Neonatal Mortality Address Similarity Score Histogram

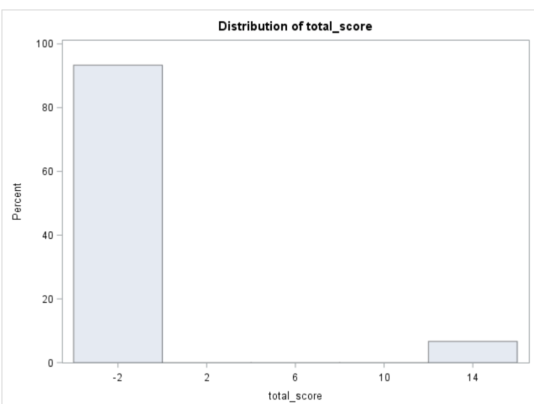


Table 11: Neonatal Mortality Age Similarity Score Histogram

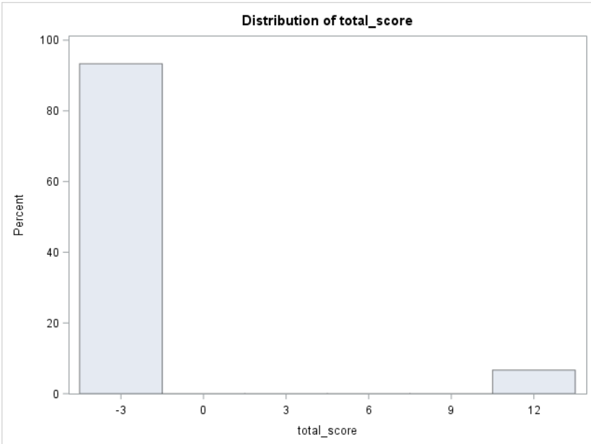


Table 12: Neonatal Mortality Sex Similarity Score Histogram

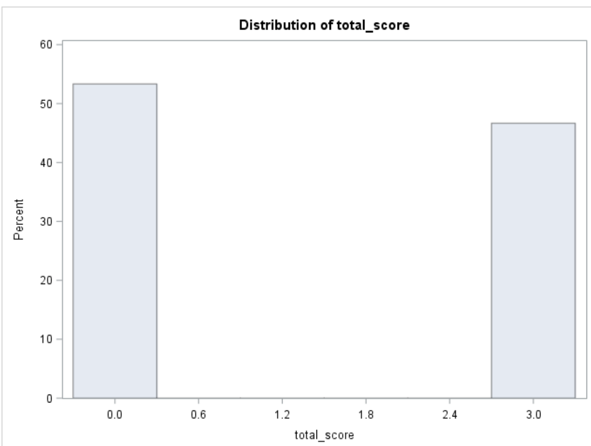


Table 13: Stillbirth Date Similarity Score Histogram

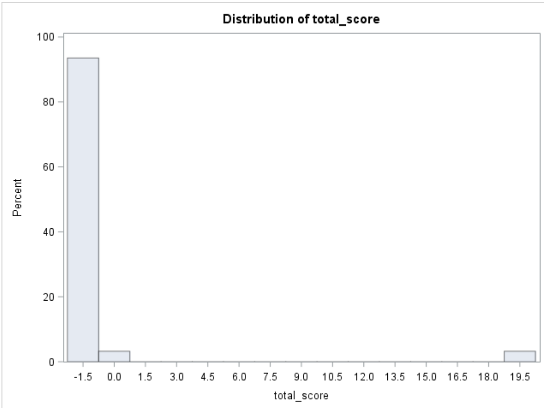


Table 14: Stillbirth Address Similarity Score Histogram

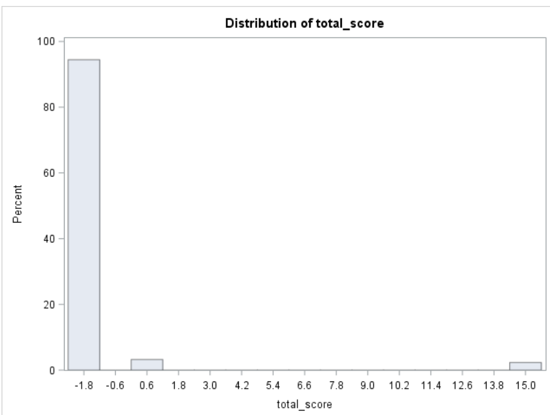


Table 15: Stillbirth Age Similarity Score Histogram

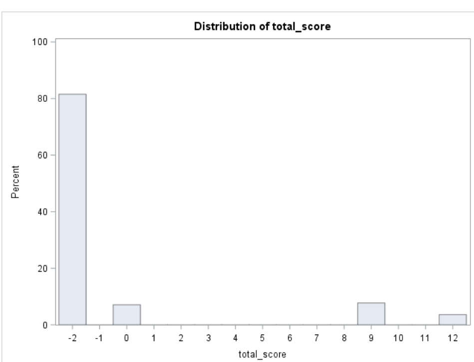


Table 16: Stillbirth Sex Similarity Score Histogram

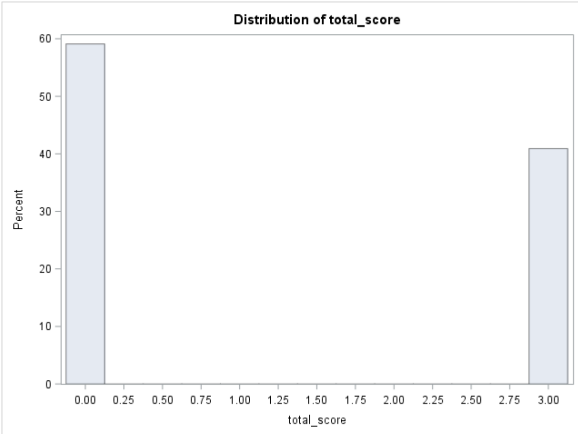


Table 17: Abortion Date Similarity Score Histogram

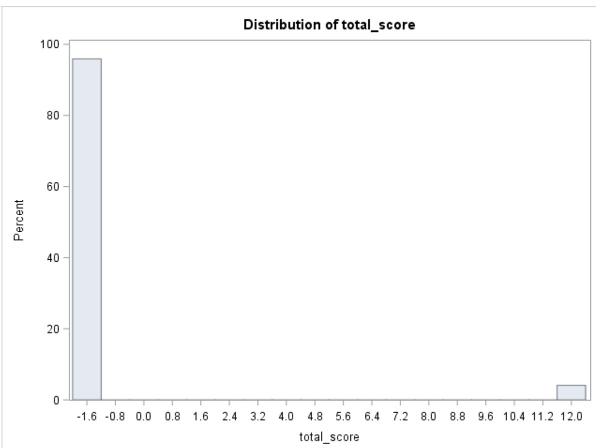


Table 18: Abortion Address Similarity Score Histogram

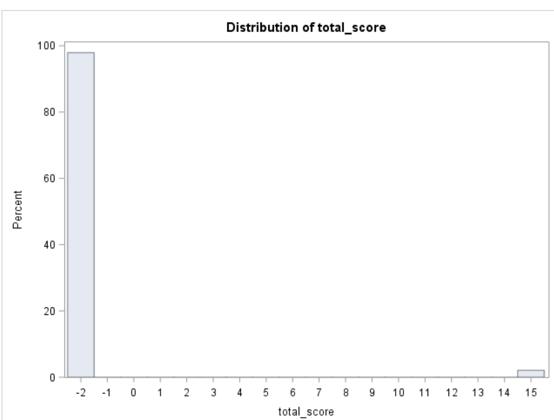


Table 19: Abortion Age Similarity Score Histogram

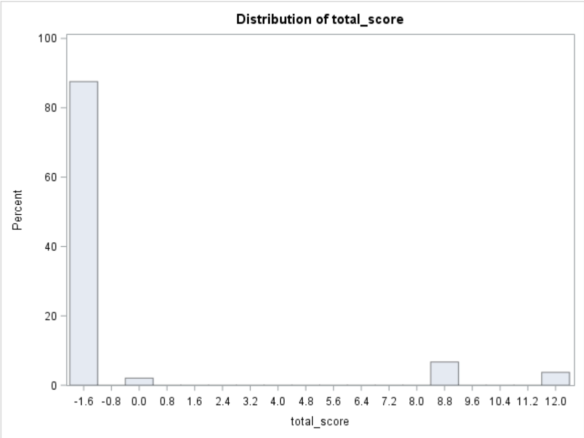
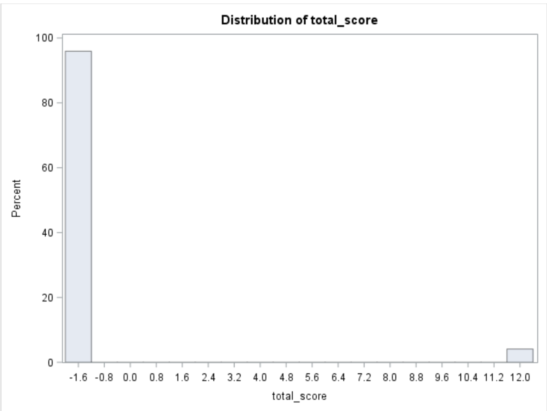


Table 20: Abortion Sex Similarity Score Histogram



SAS Code

```
**Bringing in Excel with Household Surveys**;
PROC IMPORT OUT= WORK.TEST2
            DATAFILE= "H:\CDC_ERRB\Household_Surveys_final.xlsx"
            DBMS=EXCEL REPLACE;
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;

**set up clean Household Surveys data(test2)**;
data test2a;
      length address $2;
set test2;
list=2;
```

```
obs_num=_n_;


if preg1_end111f=1 THEN livebirth=1;
if preg1_end111f=2 THEN livebirth=2;
if preg1_end111f=5 THEN livebirth=1;

if preg2_end111f=1 THEN livebirth=1;
if preg2_end111f=2 THEN livebirth=2;
if preg2_end111f=5 THEN livebirth=1;

if preg1_end111f=3 THEN stillbirth=1;
if preg1_end111f=4 THEN stillbirth=2;
if preg1_end111f=5 THEN stillbirth=1;

if preg2_end111f=3 THEN stillbirth=1;
if preg2_end111f=4 THEN stillbirth=2;
if preg2_end111f=5 THEN stillbirth=1;

if preg1_end111f=6 THEN miscarriage_abortion=1;
if preg2_end111f=6 THEN miscarriage_abortion=1;

if livedays='#VALUE!' then livedays=.;
if livedays=0 and preg1_end111f=1 THEN neonataldeath=1;
if livedays=0 and preg1_end111f=2 THEN neonataldeath=1;
if livedays=0 and preg1_end111f=5 THEN neonataldeath=1;
if livedays>0 and livedays<=28 THEN neonataldeath=1;

if age103=99 then age103=.;


if preg1_baby1_sex115='.' then preg1_baby1_sex115=' ';
if preg1_baby1_sex115='1' then preg1_baby1_sex115='M';
if preg1_baby1_sex115='2' then preg1_baby1_sex115='F';

if deathdate1='xx/xx/xxxx' then deathdate1=' ';
if deathdate1='xx/xx/2013' then deathdate1=' ';
if deathdate1='41444' then deathdate1=' ';
if deathdate1='41455' then deathdate1=' ';
if deathdate1='12/xx/2012' then deathdate1='12/01/2012';
if deathdate1='4/xx/2013' then deathdate1='04/01/2013';
if deathdate1='6/xx/2013' then deathdate1='06/01/2013';
if deathdate1='7/xx/2013' then deathdate1='07/01/2013';
if deathdate1='8/xx/2013' then deathdate1='08/01/2013';
if deathdate1='9/xx/2013' then deathdate1='09/01/2013';


deathdate1 = '12/01/2012';
deathdate99 = input(deathdate1,ddmmyy10.);
format deathdate99 ddmmyy10.;

address=propcase(address);

rename age103=age deathdate99=ddate preg1_baby1_sex115=sex;

run;
```

```
*****************************************************************************

**Bringing in excel with HIS data  - Sheet 1**;
PROC IMPORT OUT=work.test1
                  DATAFILE= "H:\CDC_ERRB\HIS1.xlsx"
                  DBMS=EXCEL REPLACE;
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;

**set up clean HIS data (test1)**;
data test1a;
      length address $2;
set test1;
list=1;
obs_num=_n_;

if Indication="ABORTION" then abortion=1;

format Date ddmmyy10.;

address=propcase(address);

rename Age=age Date=ddate;

run;

*************************************************************

**Bringing in excel with HIS data - Sheet 2**;
PROC IMPORT OUT=work.test1
                  DATAFILE= "H:\CDC_ERRB\HIS2.xlsx"
                  DBMS=EXCEL REPLACE;
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;

**set up clean HIS data (test1)**;
data test1b;
      length address $2;
set test1;
list=1;
obs_num=_n_;


if Dx="abortion" then abortion=1;
if Dx="septic abortion" then abortion=1;
if Dx="complete ab" then abortion=1;
if Dx="inevitable ab" then abortion=1;
if Dx="ab" then abortion=1;
if Dx="ABORTION" then abortion=1;
```

```
if Age=99 then Age=.;
if Age=999 then Age=.;

format Date_of_Exit ddmmyy10.;

address=propcase(address);

rename Age=age Date_of_Exit=ddate;

run;


****************************************************************************

**Bringing in excel with HIS data - Sheet 3**;
PROC IMPORT OUT=work.test1
                  DATAFILE= "H:\CDC_ERRB\HIS3.xlsx"
                  DBMS=EXCEL REPLACE;
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;

**set up clean HIS data (test1)**;
data test1c;
      length address $2;
set test1 (DROP= LMP DELIVERY_DATE COMPLICATION_TYPE);
list=1;
obs_num=_n_;

if STILLBIRTH=0 then neonataldeath=1;
if STILLBIRTH=1 then stillbirth=1;
if STILLBIRTH>2 then delete;

if SEX='FEMALE' then SEX='F';
if SEX='FEMLE' then SEX='F';
if SEX='FEMALE/FEMALE' then SEX='F';
if SEX='KE' then SEX='F';
if SEX='MALE' then SEX='M';

format DATE_ADM ddmmyy10.;

address=propcase(address);

rename AGE=age SEX=sex DATE_ADM=ddate;


run;

****************************************************************************

**Bringing in excel with HIS data - Sheet 4**;
PROC IMPORT OUT=work.test1
                  DATAFILE= "H:\CDC_ERRB\HIS4.xlsx"
```

```
                 DBMS=EXCEL REPLACE;
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;

**set up clean HIS data (test1)**;
data test1d;
      length address $2;
set test1 (DROP= LMP DELIVERY_DATE COMPLICATION_TYPE);
list=1;
obs_num=_n_;

if DELIVERY_NORMAL=1 then livebirth=1;

if Age=999 then Age=.;

format Date_Adm ddmmyy10.;

address=propcase(address);

rename Age=age Date_Adm=ddate Sex=sex;


run;

****************************************************************************
**Set final datasets - Household surveys - HIS (all 4 sheets combined)***;

DATA HIS;
      set test1a test1b test1c test1d;
RUN;

**subset each database, keeping only poor pregnancy outcome of interest, run
each poor pregnancy outcome of interest separately for the rest of the code,
otherwise it will overwrite itself**

***subset each database, keeping only neonatal deaths**;

DATA test2b;
      SET test2a;
      where neonataldeath=1;
run;

proc contents data=test2b;
run;

DATA HISb;
      SET HIS;
      where neonataldeath=1;
run;

proc contents data=HISb;
run;
```

```
***subset each database, keeping only stillbirths**;

DATA test2b;
      SET test2a;
      where stillbirth=1;
run;

proc contents data=test2b;
run;

DATA HISb;
      SET HIS;
      where stillbirth=1;
run;

proc contents data=HISb;
run;


***subset each database, keeping only abortions**;

DATA test2b;
      SET test2a;
      where miscarriage_abortion=1;
run;

proc contents data=test2b;
run;

DATA HISb;
      SET HIS;
      where abortion=1;
run;

proc contents data=HISb;
run;


**Use Proc SQL (Structured Query Language) which is similar to a data step
but takes less code to write out. Use this to create a table joining the two
data sources**;
**write each variable name as it is labeled in each source along with the
corresponding name with the other source**;
**a means it is coming from the first source (HIS) and b means it is coming
from the second source (Household Surveys)**;


**Calculating Similarity scores**;

**Following convention in the propensity matching paper, value is given if
two values agree, negative value is given if two values disagree, and no
value is given if either of the values are missing as there is no evidence
present;

**Inexact matching coding is utilized to give points for inexact matches
(misspelled names, transposed days and months in birth dates);
```

```
****************************************************************************
***************************************************************************
**Here below is code for running it one variable at a time;
**Histograms can be seen of the scores specific to each variable;


**AGE;

proc sql;
create table cartesian_join as
select

a.obs_num as obs_num_1, b.obs_num as obs_num_2,
a.age as age_1, b.age as age_2,
a.ddate as ddate_1, b.ddate as ddate_2,
a.sex as sex_1, b.sex as sex_2,
a.address as address_1, b.address as address_2,


case when a.age=. or b.age=. then 0
/*full points for an exact match*/
when a.age=b.age then 12
/*reduced points when within one year*/
when a.age = (b.age - 1) then 9
when a.age = (b.age + 1) then 9
/*negative points for disagreeement*/
else -2 end as fage_score,


calculated fage_score as total_score

from HISb as a INNER JOIN test2b as b
on b.obs_num > a.obs_num

;
quit;

proc sort data=cartesian_join;
by decending total_score;
run;

**Choosing a threshold;

**create a histogram to look and see what thresholds would be appropriate;

proc univariate data=cartesian_join;
      var total_score;
      histogram total_score;
run;


**DATE;
proc sql;
create table cartesian_join as
select
```

```
a.obs_num as obs_num_1, b.obs_num as obs_num_2,
a.age as age_1, b.age as age_2,
a.ddate as ddate_1, b.ddate as ddate_2,
a.sex as sex_1, b.sex as sex_2,
a.address as address_1, b.address as address_2,

case when a.ddate=. or b.ddate=. then 0
/*full points for an exact match*/
when a.ddate = b.ddate then 20
/*near-full points for same year, but transposed day and month*/
when year(a.ddate) = year(b.ddate) and month(a.ddate) = day(b.ddate) and
day(a.ddate) = month(b.ddate) then 12
/*reduced points when only two of the three fields match month*/
when year(a.ddate) = year(b.ddate) and month(a.ddate) = month(b.ddate) then
12
when year(a.ddate) = year(b.ddate) and day(a.ddate) = day(b.ddate) then 12
when month(a.ddate) = month(b.ddate) and day(a.ddate) = day(b.ddate) then 12
/*negative points for disagreeement*/
else -2 end as fddate_score,


calculated fddate_score as total_score

from HISb as a INNER JOIN test2b as b
on b.obs_num > a.obs_num

;
quit;

proc sort data= cartesian_join;
by decending total_score;
run;

**Choosing a threshold;

**create a histogram to look and see what thresholds would be appropriate;

proc univariate data=cartesian_join;
      var total_score;
      histogram total_score;
run;


**SEX;

proc sql;
create table cartesian_join as
select

a.obs_num as obs_num_1, b.obs_num as obs_num_2,
a.age as age_1, b.age as age_2,
a.ddate as ddate_1, b.ddate as ddate_2,
a.sex as sex_1, b.sex as sex_2,
a.address as address_1, b.address as address_2,
```

```
case when a.sex=' ' or b.sex=' ' then 0
/*full points for an exact match*/
when a.sex=b.sex then 3
/*negative points for disagreeement*/
else 0 end as fsex_score,


calculated fsex_score as total_score

from HISb as a INNER JOIN test2b as b
on b.obs_num > a.obs_num

;
quit;

proc sort data= cartesian_join;
by decending total_score;
run;

**Choosing a threshold;

**create a histogram to look and see what thresholds would be appropriate;

proc univariate data=cartesian_join;
      var total_score;
      histogram total_score;
run;


**ADDRESS;
proc sql;
create table cartesian_join as
select

a.obs_num as obs_num_1, b.obs_num as obs_num_2,
a.age as age_1, b.age as age_2,
a.ddate as ddate_1, b.ddate as ddate_2,
a.sex as sex_1, b.sex as sex_2,
a.address as address_1, b.address as address_2,

case when a.address=' ' or b.address=' ' then 0
/*full points for an exact match*/
when a.address = b.address then 15
/*negative points for disagreeement*/
else -2 end as faddress_score,


calculated faddress_score as total_score

from HISb as a INNER JOIN test2b as b
on b.obs_num > a.obs_num

;
quit;

proc sort data= cartesian_join;
by decending total_score;
```

```
run;

**Choosing a threshold;

**create a histogram to look and see what thresholds would be appropriate;

proc univariate data=cartesian_join;
      var total_score;
      histogram total_score;
run;


*****************************************************************************
*****************************************************************************
*
**Calculating Similarty Scores based off of all variables combined**
**Same methods, but total score combines all of the above variables**;

proc sql;
create table cartesian_join as
select

a.obs_num as obs_num_1, b.obs_num as obs_num_2,
a.age as age_1, b.age as age_2,
a.ddate as ddate_1, b.ddate as ddate_2,
a.sex as sex_1, b.sex as sex_2,
a.address as address_1, b.address as address_2,


case when a.age=. or b.age=. then 0
/*full points for an exact match*/
when a.age=b.age then 12
/*reduced points when within one year*/
when a.age = (b.age - 1) then 9
when a.age = (b.age + 1) then 9
/*negative points for disagreeement*/
else -7 end as fage_score,


case when a.ddate=. or b.ddate=. then 0
/*full points for an exact match*/
when a.ddate = b.ddate then 12
/*near-full points for same year, but transposed day and month*/
when year(a.ddate) = year(b.ddate) and month(a.ddate) = day(b.ddate) and
day(a.ddate) = month(b.ddate) then 11
/*reduced points when only two of the three fields match month*/
when year(a.ddate) = year(b.ddate) and month(a.ddate) = month(b.ddate) then 9
when year(a.ddate) = year(b.ddate) and day(a.ddate) = day(b.ddate) then 9
when month(a.ddate) = month(b.ddate) and day(a.ddate) = day(b.ddate) then 9
/*negative points for disagreeement*/
else -7 end as fddate_score,

case when a.sex=' ' or b.sex=' ' then 0
/*full points for an exact match*/
when a.sex=b.sex then 12
/*reduced points when within one year*/
else -7 end as fsex_score,
```

```
case when a.address=' ' or b.address=' ' then 0
/*full points for an exact match*/
when a.address = b.address then 12
/*negative points for disagreeement*/
else -7 end as faddress_score,


calculated fage_score + calculated fddate_score + calculated fsex_score +
calculated faddress_score as total_score

from HISb as a INNER JOIN test2b as b
on b.obs_num > a.obs_num

;
quit;

proc sort data= cartesian_join;
by decending total_score;
run;

**Choosing a threshold;

**create a histogram to look and see what thresholds would be appropriate;
**ODS RTF FILE='H:\CDC_ERRB\NeonateHisto.RTF';
proc univariate data=cartesian_join;
      var total_score;
      **where total_score>(-8);
      histogram total_score;
run;
**ODS RTF CLOSE;

****************************************************************************;

proc print data=cartesian_join;
      where total_score>0;
run;

proc freq data=cartesian_join;
      where total_score>0;
run;
```

IRB Approval

The Centers for Disease Control and Prevention Center for Global Health reviewed this research protocol in October 2012. The determination of "not human subjects research, primary intent is public health practice program evaluation activity" was given by the CDC CGH Office of the Associate Director of Science. Designation of "Non-Research" was given to the proposed assessment after review. The original project was determined non-research by CDC IRB and this is a secondary data analysis of that project data.

References

Barfield, W. D. (2011). "Standard terminology for fetal, infant, and perinatal deaths." Pediatrics **128**(1): 177-181.

Blencowe, H., S. Cousens, F. B. Jassir, L. Say, D. Chou, C. Mathers, D. Hogan, S. Shiekh, Z. U. Qureshi and D. You (2016). "National, regional, and worldwide estimates of stillbirth rates in 2015, with trends from 2000: a systematic analysis." The Lancet Global Health **4**(2): e98-e108.

Brennan, R. J. and R. Nandy (2001). "Complex humanitarian emergencies: a major global health challenge." Emergency Medicine **13**(2): 147-156.

Brittain, S. and D. Böhning (2009). "Estimators in capture–recapture studies with two sources." AStA Advances in Statistical Analysis **93**(1): 23-47.

Filippi, V., C. Ronsmans, O. M. Campbell, W. J. Graham, A. Mills, J. Borghi, M. Koblinsky and D. Osrin (2006). "Maternal health in poor countries: the broader context and a call for action." The Lancet **368**(9546): 1535-1541.

Fretts, R. C. (2005). "Etiology and prevention of stillbirth." American journal of obstetrics and gynecology **193**(6): 1923-1935.

Frøen, J. F., I. K. Friberg, J. E. Lawn, Z. A. Bhutta, R. C. Pattinson, E. R. Allanson, V. Flenady, E. M. McClure, L. Franco and R. L. Goldenberg (2016). "Stillbirths: progress and unfinished business." The Lancet **387**(10018): 574-586.

Glenn Wright, M. (2010). "Probabilistic Record Linkage in SAS®."

Haskew, C., P. Spiegel, B. Tomczyk, N. Cornier and H. Hering (2010). "A standardized health information system for refugee settings: rationale, challenges and the way forward." Bulletin of the World Health Organization **88**(10): 792-794.

Hook, E. B. and R. R. Regal (1995). "Capture-recapture methods in epidemiology: methods and limitations." Epidemiologic reviews **17**(2): 243-264.

Hynes, M., O. Sakani, P. Spiegel and N. Cornier (2012). "A Study of Refugee Maternal Mortality in 10 Countries, 2008—2010." International perspectives on sexual and reproductive health: 205-213.

Hynes, M., M. Sheik, H. G. Wilson and P. Spiegel (2002). "Reproductive health indicators and outcomes among refugee and internally displaced persons in postemergency phase camps." JAMA **288**(5): 595-603.

Idowu, R. T. (2014). Use of capture-recapture methodology to characterize neonatal mortality, stillbirth, and live birth rates in Farchana Refugee Camp - Chad, 2013. EIS Conferences.

Liu, L., S. Oza, D. Hogan, J. Perin, I. Rudan, J. E. Lawn, S. Cousens, C. Mathers and R. E. Black (2015). "Global, regional, and national causes of child mortality in 2000–13, with projections to inform post-2015 priorities: an updated systematic analysis." The Lancet **385**(9966): 430-440.

Nobuko, M., S. Lale, Z.-B. Sarah, I. Mie and C. Doris (2012). "A New Method for Deriving Global Estimates of Maternal Mortality." Statistics, Politics and Policy **3**(2).

O'Heir, J. (2004). "Pregnancy and Childbirth Care Following Conflict and Displacement: Care for Refugee Women in Low-Resource Settings." Journal of Midwifery & Women's Health **49**(S1): 14-18.

SAS Institute, (2016). SAS user's guide: statistics, SAS Institute Inc.

Taylor, S., B. Williams, D. Magnus, A. Goenka and N. Modi (2015). "From MDG to SDG: good news for global child health?" The Lancet **386**(10000): 1213-1214.

Toole, M. J. and R. J. Waldman (1997). "The public health aspects of complex emergencies and refugee situations 1." Annual review of public health **18**(1): 283-312.

UN (2015). Assistance to refugees, returnees and displaced persons in Africa, Report of the Secretary-General. Report of the United Nations High Commissioner for Refugees, questions relating to refugees, returnees and displaced persons and humanitarian questions. General Assembly**:** 18.

UN. (2015). "Millennium development goals." United Nations. Available online: http://www. un. org/millenniumgoals/(accessed on 1 August 2016).

UNHCR (2013). UNHCR Global Report 2012, United Republic of Tanzania. refworld.

Wardlaw, T., D. You, L. Hug, A. Amouzou and H. Newby (2014). "UNICEF Report: enormous progress in child survival but greater focus on newborns urgently needed." Reprod Health **11**(1): 82.

WHO (2012). Facts on Induced Abortion Worldwide. oneline, World Health Organization.

WHO (2013). "Global health observatory data repository: life tables." Geneva: WHO.

WHO (2016). Newborns: reducing mortality. Online, World Health Organization.

WHO (2016). Stillbirths. Maternal, newborn, child and adolescent health. online.

Wise, P. H. and G. L. Darmstadt (2015). Strategic governance: Addressing neonatal mortality in situations of political instability and weak governance. Seminars in perinatology, Elsevier.