Discussion Paper, Series A, No.2019-338

Assessing the population-wide exposure to lead pollution in Kabwe, Zambia: blood lead level estimation based on survey data

Masato Hiwatari, Daichi Yamada, Peter Hangoma,
Daiju Narita, Chrispin Mphuka, Bona Chitah, John Yabe,
Shouta MM Nakayama, Hokuto Nakata, Kennedy Choongo,
Mayumi Ishizuka

MAY.2019

Faculty of Economics and Business
Hokkaido University
Kita 9 Nishi 7, Kita-Ku, Sapporo 060-0809, JAPAN

Assessing the population-wide exposure to lead pollution in Kabwe, Zambia: blood lead level estimation based on survey data

Masato Hiwatari[1 †], Daichi Yamada[2], Peter Hangoma[3], Daiju Narita[2], Chrispin Mphuka[4], Bona Chitah[4], John Yabe[5], Shouta MM Nakayama[6], Hokuto Nakata[6], Kennedy Choongo[5], Mayumi Ishizuka[6]

[1] Faculty of Economics & Business, Hokkaido University

[2] Graduate School of Arts and Sciences, The University of Tokyo

[3] The University of Zambia, School of Public Health

[4] The University of Zambia, School of Humanity and Social Science

[5] The University of Zambia, School of Veterinary Medicine

[6] Faculty of Veterinary Medicine, Hokkaido University

Abstract

In this study, we aim to quantitatively assess the population-wide exposure to lead pollution in Kabwe, Zambia. While Kabwe is known as one of the most significant cases of environmental pollution in the world, the available information does not provide a representative figure on residents' lead poisoning conditions. To obtain a representative figure, we estimate blood lead level (BLL) of the representative sample of Kabwe by combining two datasets: BLL data collected based on residents' voluntary participation to blood sampling and socioeconomic data collected for approximately 900, randomly chosen households that represent Kabwe population. The results show that the representative mean BLL is slightly lower than the one observed in previous studies but a few times higher than the recent standard BLL of 5μg/dL above which health risks become significant.

[†] Address for correspondence: Masato Hiwatari. Faculty of Economics & Business, Hokkaido University, Kita 9, Nishi 7, Kita-ku, Sapporo 060-0809, Japan. Email: mhiwatari@econ.hokudai.ac.jp.

We are indebted to participants to our surveys in Kabwe. We are also grateful to Central Statistical Office and the Ministry of Health of Zambia, the Kabwe District Health Office, and laboratory technicians and nurses at health clinics in Kabwe. This work was conducted as a part of KAbwe Mine Pollution Amelioration Initiative (KAMPAI) project supported by JST/JICA, SATREPS (Science and Technology Research Partnership for Sustainable Development; No. JPMJSA1501).

1. Introduction

Environmental pollution is one of the largest threats to human health, particularly for the poor and vulnerable in low- and middle-income countries. Recent estimates show that, based on the Global Burden of Disease Study (GBD), pollution was responsible for 9 million deaths, with 92% occurring in low-and middle-income countries, and for 269 million disability-adjusted life-years (DALYs) (Landrigan et al. 2018). In particular, hazards of toxic chemical and heavy metal contamination, which are linked to industrial development, are contemporary and growing major problems in most of developing countries. According to Blacksmith Institute and Green Cross Switzerland (2013), nine of the top ten most polluted sites in the world were located in low- and middle-income countries.

Kabwe district, Zambia, which is one of the top ten most polluted sites and was named as the world's most toxic town by the Guardian newspaper[1], provides an illustrative example of the devastating consequences of contemporary environmental pollution in low- and middle-income countries. The town was once known for lead (Pb) and zinc (Zn) mining activities. Although the formal operation of the mine was terminated in 1994, mining residues have been left abandoned in the dumping site located close to residential areas and continue providing contamination problems for the local environment, particularly soil lead pollution (Křibek et al. 2019; Nakayama et al. 2011; Yabe et al. 2015). According to the Toxic Sites Identification Program (TSIP) Global Database[2], a global database of toxic sites in lower- and middle-income countries, Kabwe's lead contamination has the largest number of the affected population, 120,000 people, among all the

[1] The Guardian, "The world's most toxic town: the terrible legacy of Zambia's lead mines," May 28, 2017.
[2] http://www.contaminatedsites.org/TSIP/ (accessed on May 6, 2019)

confirmed (reviewed) cases of lead contamination in the world recorded on the database.

Costs of lead poisoning in Kabwe can be tremendous. Lead intake adversely affects the functioning of circulatory and nervous systems, which can be fatal at high exposure levels, whereas relatively low levels of lead exposure, including pre-natal exposure in the utero and post-natal exposure in the early infancy, can lead to a reduction in IQ and cognitive ability and cause developmental disorders (Canfield et al. 2003; Dapul and Laraque 2014; Meyer et al. 2008; WHO 2010). The adverse health effects may further result in poor school performance, reduced education attainment, behavioral disorders, and poor lifetime earnings (e.g. Aizer et al. 2018; Aizer and Currie, forthcoming; Miranda et al. 2007; Needleman 2004). The recent standards regard the blood lead level (BLL) of 5µg/dL as the reference level above which health risks become significant, and recommend an immediate chelation therapy for those with BLLs above 45µg/dL (Center for Disease Control and Prevention (CDC) 2012). In case of Kabwe, previous studies have reported that the average BLL even exceeds 45µg/dL in the neighborhood of the mining sites and that certain residents exhibit BLLs normally regarded as fatal (Caravanos, Fuller & Robinson 2014; Mbewe et al. 2015; Yabe et al. 2015, 2019).

Despite these astonishing figures, relatively little is known regarding the lead poisoning levels of the entire population in Kabwe district. Previous surveys mainly focused on children living in neighboring areas of the mine or the mine dumping site and did not provide representative estimates of lead poisoning in Kabwe. While these studies are useful for clarifying the needs for direct health interventions at pollution hotspots, the lack of pollution estimations at the district level inhibited policymakers from figuring out the exact scale of the Kabwe's pollution problem and from effectively planning macro-level policies. Moreover, in the earlier studies, as a reflection of the absence of a public mechanism for formal and compulsory blood lead testing locally, blood sampling relied on residents' voluntary participation and thus suffered from a selection bias in sampling; those voluntarily participating in the blood sampling can be those particularly caring for or suspected to have lead poisoning, who may have lower or higher BLLs than an average Kabwe resident.

The purpose of this study is to quantitatively assess the prevalence and severity of lead poisoning among the entire population in Kabwe district. We utilize two sets of data we collected in Kabwe in 2017: the BLL data from blood sampling survey and the household survey data that provide demographic, socioeconomic and geographic information of residents. Both surveys targeted the same sample of approximately 900 randomly selected households that represent the entire population in Kabwe, although the BLL data are available for a subset of the target sample owing to the residents' voluntary participation decisions. Then, we estimate the equation to predict BLLs by combining these two datasets and paying attention to the potential difference in both observable and unobservable characteristics among the participants and non-participants. Finally, we present our estimates on the representative BLLs, including those not participating in blood sampling.

The contribution of our study is twofold. First, we contribute to the studies that quantify the global burden of general disease or pollution-related disease (e.g., GBD; TSIP; Landrigan and Fuller 2012; Landrigan et al. 2018; Prüss-Ustün et al. 2011). Despite the substantial impacts of lead poisoning on the global disease burden, many gaps in knowledge remain and the existing data do not provide the exact burden of disease at contaminated sites (Landrigan et al. 2018). Our study is the first attempt to obtain representative estimates of lead exposure of residents in the entire

Kabwe District, which is regarded as one of the most contaminated areas in the world today. It provides useful information for policy and public health researchers and practitioners interested in the global burden of chemical pollutants.

The second contribution is about the unique data collection method that we employed to obtain a representative figure of lead poisoning. The individual data collected from the clinic- and hospital-based surveys could suffer from the selection bias of respondents as investigators cannot perfectly control the willingness of the participation of potential respondents (e.g., Gluud 2006; Ness et al. 2009; Tripepi et al. 2010). We attempt to address this selection bias problem by conducting two surveys, the blood sampling survey to measure BLLs and the socioeconomic survey for a representative sample of population to measure both factors affecting BLLs and participation decisions, under a jointly coordinated framework. Such a methodology would not only be helpful in mitigating the bias in case of lead pollution in Kabwe, but also be applicable to other cases where formal and compulsory testing for a disease is absent.

The remainder of this paper is structured as follows. Section 2 describes our data collection methods and potential concerns on the BLL data. Section 3 presents the methodology to estimate BLLs. Section 4 provides the estimated equation and Section 5 demonstrates our final estimates on the representative BLLs. Section 6 concludes.

2. Data description

2.1 Data collection

We use the data of two surveys jointly conducted in July to September 2017 in Kabwe as parts of

KAMPAI project: the Kabwe Household Socioeconomic Survey (KHSS) 2017 conducted by the Central Statistical Office of Zambia (CSO) under supervision of the economics team of the authors, and BLL data collected by the medical team of the authors. The two surveys target the same sample households randomly chosen in two steps. We chose 40 out of 384 Standard Enumeration Areas (SEAs) in Kabwe defined by Central Statistical Office of Zambia in the first step, and then we chose 25 households (and a few replacement households) from each chosen SEA in the second step, so that the sample households are representative of whole Kabwe district. For details of the sampling methods, see Hiwatari et al. (2018).

The KHSS 2017 conducted interview surveys with 895 households (4,900 individuals) and collected information on individuals' and households' demographic and socioeconomic statuses and activities. The response rate was fairly high with 88.2%.

The BLL data were collected by the medical team of the authors. Alongside the KHSS 2017, we invited up to four persons (two children and their parents or guardian) from each sample household to blood sampling. 11 local clinics were chosen as the venue. Because it was not feasible to allocate survey technicians to 11 clinics simultaneously, we formed four groups of technicians and visited clinics sequentially. On average, we stayed in a clinic for two weeks. Also, to avoid overcrowding, we sequentially invited households. That is, we asked all households in a SEA to attend clinics during the assigned dates, which typically have a three-day window from the next day of invitation, although we tried to be flexible and conducted blood sampling even if some individuals attend clinics outside of the assigned dates, including Saturday and Sunday, as long as technicians had not moved to the next clinics. Unlike the KHSS 2017, however, the participation rate was not high. 827 sample individuals (16.9%) covered in the KHSS 2017 participated in blood sampling.

For the participants to blood sampling, their BLL were measured using kit LeadCare II©. Because the measurable range of LeadCare II© is 3.3-65μg/dL, the BLLs above 65μg/dL were measured after diluting blood and those below 3.3μg/dL were replaced by 2.6μg/dL following the results in the previous studies conducted in Kabwe.[3] For operational details, see Yabe et al. (2019).

2.2 Descriptive statistics and potential concerns of observed BLLs

Table 1 shows the mean BLLs, using the data of 827 sample individuals participated in blood sampling. Generally, the observed BLLs are high, with the sample mean being 4 times greater than the reference level of 5μg/dL. Male appears to have higher BLL than female. The mean BLLs for children aged 0-5 years and 6-10 years are similar. Adults tend to have lower BLLs but the mean still exceeds the reference level. In addition, Yabe et al. (2019) show that the variation of BLL is also large and that some individuals have even fatally high BLLs, particularly in neighbourhood of the mine dumping site.

[Table 1 here]

However, the observed BLLs may fail to represent the lead poisoning condition of the whole Kabwe population. The participation rate may not be low for a survey that conducts blood sampling on the voluntary basis. However, owing to endogenous participation decisions, the characteristics of the participants apparently deviate from those of the population. Figure 1

[3] In another study conducted in Kabwe by the authors, blood samples containing less than 3.3μg/dL of lead were re-examined using ICP-MS and the mean BLL was 2.6μg/dL.

demonstrates the participation rate by age group. The rate is 20 to 30 percent among young children and adult women whereas adult men have lower participation rates and the rate is almost zero among adolescents. The participation rate difference disturbs the age distribution among the participants. Furthermore, the participants and non-participants have different socioeconomic characteristics (Table 2). Among children, participants tend to be in small households and their mothers tend to have less education. Among adults, the difference in characteristics is even clearer than among children. Participants tend to be younger, more likely to be female, less educated, and in poorer and smaller households than non-participants. Furthermore, the participants may have certain unobservable characteristics that lead to lower or higher BLLs. For example, those who are cautious about lead poisoning may be willing to participate in blood sampling while having low BLLs. Conversely, individuals with innate abilities leading to low BLL, such as a low rate of lead absorption, may not participate in blood sampling because he/she does not feel any symptom of lead poisoning. Considering these possibilities, while the geographic and age coverages of our BLL data are wider than the previous surveys focusing only on children residing in the neighborhood of the mine, the observed BLL data of ours can still fail to represent the lead poisoning condition of the whole Kabwe population.

[Figure 1 here]

[Table 2 here]

3. Estimating BLLs: methodology

3.1 The model of BLL

In order to obtain representative figures of the BLLs of Kabwe residents, differences in observed and unobserved characteristics between participants and non-participants need to be adequately controlled. To do this, we first estimate the equations to explain BLL, separately for children aged 0-10 years and adults aged 19-69 years, using data of the participants. Then, we predict BLLs for all individuals aged 0-69 years, including non-participants, using the estimated equations.

Generally, BLL depends on the ambient pollution, the opportunities of exposure to pollution, the biophysiology of lead absorption and excretion, and the knowledge and technologies to avoid lead poisoning. We assume that these are explicable from one's own and household's characteristics collected in the KHSS 2017, which are available regardless of participation to blood sampling. Specifically, we assume the following function of BLL:

$$\log BLL_i = \beta_1 \log distance_i + f^{dir}(direction_i) + \beta_2 altitude_i + f^{age}(age_i) + \boldsymbol{X}_i\boldsymbol{\gamma}' + \varepsilon_i, \quad (1)$$

We measure the ambient pollution levels by the distance, direction and altitude of one's household location, where the former two are relative to the mine dumping site. Biophysiological and behavioural factors are measured by age and $\boldsymbol{X}_i$, various individual and household characteristics. ε_i is the error term that captures, for example, BLL's casual fluctuations, the literal measurement errors of BLL, and the effects of unobservable factors such as innate biophysiological abilities. Although we provide only one equation above, we assume different equations for children and adults because the factors affecting BLL could be different. Below, we discuss the details of the functional specification.

3.1.1 Distance, direction and altitude

The literature has pointed out that lead is delivered from the mine or its dumping site to the surrounding areas by wind and water flows (Křibek et al. 2019; Nakayama et al. 2011; Tembo et al. 2006). Consequently, the distance from these sites is negatively correlated with the ambient lead level. We define $distance_i$ as the distance between the mine dumping site and the location of i's household and expect $\beta_1 < 0$.

The direction and altitude also play roles in predicting ambient lead pollution. The literature generally finds that the soil lead contamination spreads to the western side of the mine dumping site, suggesting the role of the prevailing wind tending west (Křibek et al. 2019; Tembo et al. 2006). The soil lead contamination is also slightly extended to the eastern side, along the canal tending south-east (Nakayama et al. 2011; Water Management Consultants Ltd. 2006). The northern to northeastern side is least contaminated, possibly because of high altitude in these areas.

We hypothesize that, after controlling the distance, lead pollution is the greatest in the western side of the mine dumping site. We define $direction_i$ as the radian of the acute angle between the line from the mine dumping site to due west and that to i's household. $direction_i$ being close to zero implies that the location of the household is west to the mine dumping site, being close to $\pi/2$ implies that the location is either north or south, and being close to π implies that the location is east. We examine two specifications for the functional form of $f^{dir}(direction_i)$: one with $direction_i$ alone and the other with $direction_i$ and $direction_i^2$. The former specification assumes that lead pollution constantly diminishes as the household location moves from the west to the east whereas the latter allows more flexibility, such as the eastern side being more contaminated than in the northern and southern sides. We also use the altitude in meter to directly control the effects of land elevation although its effect could be absorbed by the direction variables to some extent.

3.1.2 Age

Generally, BLL is largely affected by age. Young children are at high risk of lead poisoning. Playing outside in soil and mud and children's age-appropriate hand-to-mouth behaviours expose them to lead, and their gastrointestinal absorption of lead is higher than adults' (WHO 2010). Fetuses and infants receive lead from mothers in utero and through breastfeeding if mothers are highly exposed to lead (CDC 2010). Consequently, BLL often reaches the peak at or before the age of 24 months and decreases as children grow up (CDC 1997; Canfield et al. 2003; Dietrich et al. 1992).

Following these views, we assume a non-linear relationship between children's age and BLL and define the functional form of $f^{age}(age_i)$ for children as

$$f^{age}(age_i) = I(age_i < 2) \times \left[\phi_0 + \phi_1 mage_i + \phi_2 mage_i^2\right] + \phi I(age_i \geq 2) \times age_i, \quad (2)$$

where $mage_i$ is the monthly age and $I(\cdot)$ is the indicator function taking one if the argument condition is satisfied. That is, we assume an inverted-U relationship between the logarithmic BLL and age up to 23 months old but a constantly decreasing one for children aged 2 years or above.

For adults, there is no clear biophysiological foundation for the age-BLL relationship. The literature points out that BLL of women can change during the periods of pregnancy and lactation, although it can both increase and decrease (CDC 2010; Hertz-Picciotto et al. 2000; Rothenberg et al. 1994). To flexibly capture their BLL-age relationship, we categorize adults by age groups (19-29 years, 30-39 years, 40-49 years, 50-59 years, and 60-69 years) and use their interaction terms with gender.

3.1.3 Other covariates

For children, we use the following individual and household characteristics as $\boldsymbol{X}_i$: the dummy for female; mother's education level expressed in education years that measures households' knowledge of lead pollution and poisoning[4]; the dummy taking 1 if mother is absent; the dummy for the gender of household head taking 1 if the head is female; the household size; the log of per capita household expenditure that measures the living standard. We also use the locational fixed effects by categorizing the household location into four at the SEA level; the urban area, small-scale farming area, large-scale farming area, and Makululu compound. While the entire Kabwe district is officially categorized as urban under national standards, the landscapes and main economic activities are diverse even within Kabwe district. Makululu compound is a poor area where some households reside without formal registration and public services are poorly delivered.

For adults, we keep using household size and the household location fixed effects. Instead of mother's education, here their own education is used. We also use the dummy for marital status giving one if an individual is either married or co-habiting. In addition, we use the length of residence in Kabwe expressed in years to account for the effect of long-term exposure to lead although such an effect could be weak either because of its correlation with age. Meanwhile, the per capita household expenditure is not used here because, unlike for children, this factor is not exogeneous for adults. The mother's absence and head's gender are not used, either.

[4] The survey asked the exact grade to those with secondary education or less but not to those with tertiary education. The grade of those with tertiary education is replaced by 16 years here. If mother is absent, then mother's education years is replaced by zero. The effect of mother being absent and that of having mothers with no education are distinguished by the dummy for absent mothers.

3.2 Estimation method

To estimate (1) and obtain the predicted BLL, we consider two estimation methods. The first and simpler one is OLS. If the difference in the characteristics of participants and non-participants is fully controlled by independent variables, then OLS provides the unbiased estimate of (1) that is applicable for the whole sample. However, as is discussed in Section 2.2, unobservable characteristics can affect both BLL and the decisions to participate in blood sampling, and this can bias the OLS estimate of equation (1).

Although our results show that such bias is negligible, we check it by Heckman's sample selection model. We estimate equation (1) simultaneously with the following selection equation:

$$\Pr(i\ participates) = \Phi(\delta_1 \log distance_i + g^{dir}(direction_i) \quad (3)$$
$$+ g^{age}(age_i) + \delta_2 altitude + \boldsymbol{X}_i\boldsymbol{\xi}' + \zeta window_i).$$

For both children and adults, the distance, direction, altitude and $\boldsymbol{X}_i$ appeared in (1) will continue being used. The form for $g^{dir}(direction_i)$ follows the one in equation (1), either with or without $direction_i^2$. Regarding the effect of age, we continue using the interaction of the age categories and gender for adults. For children, because the participation rate reaches the peak at the ages of 2-4 years and then constantly decreases in age, albeit with fluctuations, we assume the following functional form:

$$g^{age}(age_i) = \theta_0 I(age_i = 0) + \theta_1 I(age_i = 1) + \theta age_i. \quad (3)$$

As an exclusion restriction, we use the days of the effective window of blood sampling referred to by $window_i$. As is mentioned earlier, while we sequentially invited individuals to blood sampling with assigning the dates, we also conducted blood sampling for those who attended clinics outside the assigned window as long as the clinics were available for blood sampling.

Consequently, the window for blood sampling is effectively the number of days between the date of the initial invitation and the final day of blood sampling in the assigned clinics.

The effective window provides a conditionally independent variation to the probability of participation. It varies by SEA, and those residing in SEAs invited early or assigned Mon-Tue-Fri windows had longer effective windows for blood sampling and would more easily manage to attend clinics than those residing in SEAs invited late or assigned Wed-Thu-Fri windows. The order of the invitation is out of the control of individuals. It could be pointed out that the effective window could be to some extent correlated to the geographic characteristics of the household location. That is, a clinic in a densely populated area tend to cover more households and operate blood sampling longer than that in a sparsely populated area and, consequently, those residing in densely populated area tend to have longer effective windows. Nevertheless, such a correlation is controllable by the locational dummies and other independent variables such as distance and direction. In this sense, the effective window can conditionally serve as an exclusion restriction.

4. Results on BLL equation

4.1 Children

Table 3 shows the estimation results for BLL of children aged 0-10 years. Columns (I) and (II) show the OLS estimates, with the difference in the functional form of $f^{dir}(direction_i)$. In either estimation, distance and direction have strongly significant coefficients. The explanatory powers of these factors are so large that R^2 remains above 0.7 even if other independent variables are dropped and that R^2 decreases to 0.4 by dropping these factors. Considering such strong powers of these factors, and because the values of these variables are similar for households living in neighborhood, we hereafter cluster standard errors at the SEA level. Regarding the coefficient of distance, its volume is close to unity below zero, implying that BLL is inversely proportional to distance. Regarding $f^{dir}(direction_i)$, $direction_i$ has a negative coefficient in (I), capturing the feature that the western side of the mine dumping site is the most polluted. However, in column (II), both $direction_i$ and $direction_i^2$ have significant coefficients and R^2 further improves. The relationship between BLL and direction is U-shaped with the bottom peak lying around $direction_i = 3\pi/4$, the northeast and southeast of the mine dumping site (the sample does not cover the southeastern areas, however). The altitude does not have a significant coefficient.

Age is also a factor determining BLL. The quadratic form for children aged 0-23 months show that BLL reaches the peak at the age of 15 months. Interestingly, this corresponds to the timing to stop breastfeeding. On average in Kabwe, children are breastfed up to 15.8 months. This suggests the critical role of lead transfer through breastfeeding. From the age of 2 years, BLL constantly decreases as children grow up, with a one-year increase of age reducing BLL by approximately 5%.

Among other factors, mother's education reduces BLL significantly if $direction_i^2$ is not used, but this effect becomes insignificant if $direction_i^2$ is added. Meanwhile, the per capita household expenditure does not have a significant effect. These results suggest that children's BLLs are primarily determined by the level of ambient lead and their own biophysiological factors and that mothers' knowledge of lead and living standards do not affect children's BLL.

To check sample selection bias, we move to the results under Heckman's sample selection model. Column (III) shows the selection equation under probit, and column (IV) shows the estimates of

BLL, in both which $direction_i^2$ is not used. Columns (V) and (VI) show the results with $direction_i^2$. In either model, the probability of participation depends on age, the household size, and the direction of the household location. The effective window raises the probability of participation significantly. However, other factors have limited explanatory power, which is indeed intuitive since the characteristics of participants and non-participants do not greatly differ among children. In both models, the resulting estimates of equation (1) are similar to the ones under OLS. The inverse Mills ratio does not have a significant effect. In addition, the Wald test cannot reject the independence of equations in either model. Therefore, the participation bias is limited for children in terms of unobservable factors and OLS results are sufficient for predicting BLL of non-participants.

4.2 Adults

Then we move to estimation of adult BLL. Columns (I) and (II) in Table 4 show the OLS estimates without and with $direction_i^2$. Distance and direction have similar effects to those for children. The coefficient of the logarithmic distance is close to unity below zero. $direction_i$ has a significant and negative coefficient in column (I) but both $direction_i$ and $direction_i^2$ are strongly significant in column (II) with BLL reaching the bottom peak in the northeast or southeast of the mine dumping site. Unlike for children, altitude has a significantly negative effect if $direction_i^2$ is used. The sign is intuitive since lead-containing dust and water would flow less to highly elevated areas than to lowly elevated areas.

Age and gender do not appear to have systematic effects but women aged 40-49 and 60-69 years have lower BLL than men in the same age groups. Own education has a significantly negative effect on BLL. Unlike for children, knowledge of lead or living standards predictable from their education appear to affect adult BLL, possibly because they can control their own behaviours but controlling children's behaviours fully is difficult. Marital status, the length of residence in Kabwe, and household size do not have a significant coefficient.

Then columns (III) to (VI) show the results in Heckman's sample selection model. The participation decisions of adults appear to depend more on individual and household characteristics than those of children. Age-gender interaction terms fit the participation tendency shown in Figure 1. Highly educated individuals participate less likely, married individuals participate more likely, and those in large households participate less likely. The effective window significantly raises the probability of participation. However, similarly to the results for children, the inverse Mills ratio does not have a significant effect. The Wald test statistic cannot reject the independence of the two equations.

4.3 Predicting BLLs

The results so far show that BLL is largely predictable from the observable characteristics. Distance and direction have particularly significant effects on BLL whereas the effect of age for children is significant and consistent to their biophysiological and behavioural development. Regarding the functional form for direction, while both the specification with $direction_i$ alone and the quadratic specification fit well, the latter specification provides better fits and consistent results. Sample selection bias due to unobservable factors is not prominent so that we can rely on OLS estimates.

Considering these features, we employ OLS estimates with the quadratic form of direction, the ones in column (II) of Tables 3 and 4, to predict BLL of 2,528 children aged 0-18 years and 2,202 adults aged 19-69 years.

To predict BLL, however, an additional potential caution needs to be noted. That is, we predict BLLs of children aged 11-18 years from the estimated equation for children aged 0-10 years by extending the declining trend of BLL in age. It would be reasonable to expect BLL to continue decreasing after the age of 10 because their physical and behavioural development would continue their lead exposure. Nevertheless, whether that log-linear declining trend of BLL remains holding up to the age of 18 years is not clear. For example, the decline may stop or slow down before the age of 18 years when the body growth stops or slows down. With the limited sample size of children aged 11-18 years, there is no *a priori* way to examine how BLL of children evolves after the age of 10 years. The previous studies, including ones for Zambia and other countries, generally do not provide information on the age-BLL relationship for adolescents, either. Instead, let us examine the appropriateness of applying the estimated equation to children aged 11-18 years *a posteriori* with results.

5. Representative figure of lead poisoning

Now we examine the current situation of lead poisoning in Kabwe using the predicted BLLs. Table 5 shows the mean predicted BLLs and the percentages of those with BLLs lower than 5μg/dL, greater than 5μg/dL but lower than 45μg/dL, and greater than 45μg/dL. To calculate these values, the sampling weights are used so that the values are representative of the whole Kabwe district. Among the whole sample, the mean BLL of the Kabwe residents is predicted to be 15.42μg/dL. This level is lower than the mean of observed BLL among participants shown in Table 1, implying that participants are those with higher BLL than a representative individual in

Kabwe, but is still three times higher than the reference level of 5μg/dL employed in developed countries. The difference of the mean BLL by gender is minor and smaller than that of the observed mean BLL. The remaining columns show the information for five age groups. BLL tend to decrease in age among children. An astonishing feature is that only 6.0% of children aged 0-5 years and 10.7% of children aged 6-10 years have BLL below 5μg/dL although we do not solely pay attention to the neighbourhood of the mine dumping site. The mean BLL of children aged 11-18 years fits in the middle of that of younger cohort and that of adults aged 19-39 years.

[Table 5 here]

Figure 2 depicts the relationship of the predicted BLL and age in more detail, the mean by the solid line and the upper and lower borders of 95% confidence interval by the dotted ones. After reaching the peak at the age of 12-17 months, BLL demonstrates a declining trend in age albeit with fluctuations. An important point to be noted is that BLLs of children aged 18 years and adults aged 19-29 years are smoothly connected, rather than making a jump. This suggests that extending the negative trend of BLL in age up to the age of 18 years provides reasonable estimates of BLL for children aged 11-18 years.

[Figure 2 here]

6 Discussion and conclusion

Combining the BLL and socioeconomic data jointly collected, we aimed to obtain a representative figure of the current lead poisoning condition in Kabwe. Because of the absence of a formal and compulsory blood testing, researchers, including us, can obtain BLL data only from individuals voluntary participating to blood sampling but it often turns out difficult to obtain representative BLL data. We investigated to overcome this issue by conducting socioeconomic survey for a representative sample of the entire Kabwe population, by estimating the BLL equation, and by predicting BLL of the representative sample of individuals, including those not participating in blood sampling.

The results clarify the extent to which the residents in Kabwe are exposed to lead, which would provide a basis for the discussion on the scale of lead pollution and the cost-benefit analysis of potential remediation strategies. On one hand, the representative mean BLL is lower than the observed mean BLL of the participants to blood sampling. Young children are overrepresented among the participants compared to older children whereas young children tend to have high BLLs. Highly educated adults, who tend to have low BLLs, are underrepresented compared to lowly educated ones. These factors appear to contribute to the upward bias. On the other hand, the representative mean BLL is still 2-6 times higher than the reference BLL of 5μg/dL and 76.8% of residents have BLLs above that reference level. That proportion implies that 155,000 to 207,000 individuals[5] have BLLs suspected of lead poisoning, more than 120,000 individuals in the TSIP Global Database, for example.

The results also provide implications for potential interventions. Strong explanatory powers of the geographic characteristics of the household's location, such as the distance and direction from the mine dumping site, suggest that individuals are exposed to lead at or around home delivered from the mine dumping site by wind and water flow. High BLLs are observed in the

[5] The figure varies by population estimates. The population was 202,360 in census in 2010 and is 270,389 in our household survey estimate.

neighbourhood of the mine dumping site and, therefore, medical interventions such as chelation therapy can focus on such areas rather than on the whole Kabwe district. However, interventions to reduce of lead delivery would be of fundamental importance as they will reduce sources of lead exposure and benefit individuals whose BLLs are not as high as chelation therapy is applicable. Recently, fences have been built around the mine dumping site to prevent scavenging. While this could reduce the events of acute lead poisoning, it would not be sufficient for reducing lead poisoning in the whole Kabwe district. Meanwhile, the insignificant effects of mothers' education and household income on children's BLLs suggest that children are exposed to lead regardless of parental knowledge on lead poisoning and living standards. The extent to which parents prevent their children from lead exposure, either intentionally or unintentionally, appears limited and, in this sense, policy interventions could reduce lead poisoning without crowding out self-protection behaviours greatly, at least for children.

Finally, our results show a virtue of collecting information for those not participating in blood sampling or other forms of medical testing. In developing countries without formal and compulsory medical testing, testing based on voluntary participation would face the problem of participation bias, which could bias the burden of a disease either positively or negatively and prevent efficient execution of policy intervention.

Reference

Aizer A, Currie J, Simon P, and Vivier P. 2018. Do Low Levels of Blood Lead Reduce Children's Future Test Scores? *American Economic Journal: Applied Economics* **10**: 1–36.

Aizer A and Currie J. (Forthcoming). Lead and Juvenile Delinquency: New Evidence from Linked Birth, School and Juvenile Detention Records. *Review of Economics and Statistics*.

Blacksmith Institute and Green Cross Switzerland. 2013. *The Worlds Worst 2013: Top Ten Toxic Threats*. Retrieved from Pure Earth, http://www.worstpolluted.org/. Accessed on May 9

2019.

Bose-O'Reilly S, Yabe J, Makumba J, et al. 2018. Lead intoxicated children in Kabwe, Zambia. *Environmental Research* **165**: 420-424.

Canfield RL, Henderson CR, Cory-Slechta DA, et al. 2003. Intellectual Impairment in Children with Blood Lead Concentrations below 10 µg per Deciliter. *The New England Journal of Medicine* **348**: 1517-1526.

Caravanos J, Fuller R, and Robinson S. 2014. Severe Environmental Contamination and Elevated Blood Lead Levels Among Children – Zambia, 2014. *Morbidity and Mortality Weekly Report*, **63**: 1013.

Centers for Disease Control and Prevention (CDC). 1997. *Screening Young Children for Lead Poisoning: Guidance for State and Local Public Health Officials*. Atlanta: CDC.

—————. 2010: *Guidelines for the Identification and Management of Lead Exposure in Pregnant and Lactating Women. Atlanta*: CDC.

—————. 2012. *Low level lead exposure harms children: a renewed call for primary prevention. Report of the Advisory Committee on Childhood Lead Poisoning Prevention of the Centers for Disease Control and Prevention*. Atlanta: CDC.

Dapul H, and Laraque D. 2014. Lead Poisoning in Children. *Advances in Pediatrics* **61**: 313-333.

Dietrich KN, Berger OG, Succop PA, et al. 1993. The Developmental Consequences of Low to Moderate Prenatal and Postnatal Lead Exposure: Intellectual Attainment in the Cincinnati Lead Study Cohort Following School Entry. *Neurotoxocology and Teratology* **15**: 37-44.

Global Burden of Disease Study (GBD). The Lancet. https://www.thelancet.com/gbd. Accessed on May 17, 2019.

Gluud LL. 2006. Bias in Clinical Intervention Research, *American Journal of Epidemiology* **163**: 495-501

Hertz-Picciotto I, Schramm M, Watt-Morse M, et al. 2000. Patterns and Determinants of Blood Lead During Pregnancy. *American Journal of Epidemiology* **152**: 829-837.

Hiwatari M, Yamada D, Hangoma P, et al. 2018. *Kabwe Household Socioeconomic Survey (KHSS) 2017 Report*. Sapporo, Japan: Shobunsha. Also available at KAbwe Mine Pollution Amelioration Initiative, http://satreps-kampai.vetmed.hokudai.ac.jp/publications/.

Kříbek B, Nyambe I, Majer V, et al. 2019. Soil contamination near the Kabwe Pb-Zn smelter in Zambia: Environmental impacts and remediation measures proposal. *Journal of Geochemical Exploration* **197**: 159-173.

Landrigan PJ, and Fuller R. 2012. Environmental pollution: an enormous and invisible burden on health systems in low-and middle-income counties. World Hospitals and Health Services. **50**(4): 35-40.

Landrigan PJ, Fuller R, Acosta NJR, et al. 2018. The Lancet Commission on Pollution and Health. *The Lancet* **391**: 462-512

Mbewe A, Muneku E, Michelo C, and Siulapwa Y. 2015. Blood Lead Levels and Likely Adverse Health Outcomes in Under Five Children in Kabwe, Zambia. *International Journal of Sciences: Basic and Applied Research* **24**(5): 171-183.

Meyer PA, Brown MJ, and Falk H. 2008. Global approach to reducing lead exposure and poisoning. *Mutation Research* **659**: 166-175.

Miranda ML, Kim D, Galeano MA, Paul CJ, Hull AP, Morgan SP. 2007. The relationship between early childhood blood lead levels and performance on end-of-grade tests. *Environmental Health Perspectives* **115**:1242-1247.

Nakayama SMM, Ikenaka Y, Hamada K, et al. 2011. Metal and metalloid contamination in roadside soil and wild rats around a Pb–Zn mine in Kabwe, Zambia. *Environmental Pollution* **159**: 175-181.

Needleman H. 2004. Lead Poisoning. Annual Review of Medicine **55**: 209-222.

Ness KK, Leisenring W, Goodman P, et al. 2009. Assessment of Selection Bias in Clinic-Based Populations of Childhood Cancer Survivors: A Report from the Childhood Cancer Survivor Study. *Pediatric Blood & Cancer* **52**: 379–386.

Prüss-Ustün A, Vickers C, Haefliger P, Bertollini R. 2011. Knowns and unknowns on burden of disease due to chemicals: a systematic review. *Environmental Health* **10**(9):1-15.

Rothenberg SJ, Karchmer S, Schnaas L, et al. 1994. Changes in Serial Blood Lead Levels During Pregnancy. *Environmental Health Perspectives* **102**: 876-880.

Tembo BD, Sichilongo K and Cernak J. 2006. Distribution of copper, lead, cadmium and zinc concentrations in soils around Kabwe town in Zambia. *Chemosphere* **63**: 497-501.

Toxic Sites Identification Program (TSIP) Global Database. Blacksmith Institute. http://www.contaminatedsites.org/TSIP/. Accessed on May 17, 2019.

Tripepi G, Jager KJ, Dekker FW, and Zoccali C. 2010. Selection Bias and Information Bias in Clinical Research. *Nephron Clinical Practice* **115**: c94–c99.

Water Management Consultants, ZCCM Investments Holdings PLC, and Copperbelt Environment Project. 2006. Kabwe Scoping & Design Study, Project Synthesis:1-5.

WHO. 2010. *Childhood lead poisoning*. Geneva: WHO Press.

Yabe J, Nakayama SMM, Ikenaka Y, et al. 2015. Lead poisoning in children from townships in the vicinity of a lead–zinc mine in Kabwe, Zambia. *Chemosphere* **119**: 941-947.

Yabe J, Nakayama SMM, Nakata H, et al. 2019. *Current Trends of Blood Lead Levels, Distribution Patterns and Exposure Variations among Household Members in Kabwe, Zambia*. Mimeograph.

Table 1 Descriptive statistics of observed BLLs

	All sample	Male	Female	Age 0-5 years	Age 6-10 years	Age 11-18 years	Age 19-39 years	Age 40-69 years
Mean	21.89	23.78	20.43	29.06	29.50	23.52	13.77	18.97
Observations	827	361	466	201	154	21	233	195

Table 2 Difference in socioeconomic characteristics of the participants and non-participants

	Participants' mean	Non-participant's mean	Difference	Std. err.
Among children aged 0-10 years				
Age	4.946	5.272	-0.325*	0.190
Female	0.465	0.491	0.0262	0.0303
Household size	6.414	7.002	-0.588***	0.163
Mother's education				
None or mother not in household	0.223	0.266	-0.0432	0.0265
Lower primary	0.0676	0.587	0.0089	0.0145
Higher primary	0.301	0.237	0.0641**	0.0263
Lower secondary	0.228	0.215	0.0133	0.0251
Higher secondary	0.141	0.160	-0.0196	0.0220
Tertiary	0.0394	0.0630	-0.0235*	0.0141
Log household expenditure per capita	5.693	5.758	-0.0641	0.0538
Observations	355	1,159		
Among adults aged 19-69 years				
Age	39.87	34.13	5.736***	0.678
Female	0.650	0.506	0.143***	0.0266
Household size	5.694	6.483	-0.789***	0.152
Own education				
None	0.0607	0.0828	-0.0221	0.0144
Lower primary	0.0911	0.0414	0.0497***	0.0117
Higher primary	0.308	0.192	0.116***	0.0219
Lower secondary	0.273	0.219	0.0544**	0.0225
Higher secondary	0.182	0.295	-0.113***	0.0238
Tertiary	0.0841	0.170	-0.0859***	0.0193
Log household expenditure per capita	5.801	6.095	-0.295***	0.0497
Observations	428	1,859		

* p<0.1; ** p<0.05; *** p<0.01.

Table 3 Estimation results of BLL of children aged 0-10 years

	(I) OLS	(II) OLS	(III) 1st stage Heckman	(IV) 2nd stage Heckman	(V) 1st stage Heckman	(VI) 2nd stage Heckman
Log distance	-0.981***	-0.987***	-0.0481	-0.973***	-0.0378	-0.975***
	(0.0498)	(0.0622)	(0.0548)	(0.0476)	(0.0609)	(0.0572)
Direction	-0.304***	-1.162***	-0.173*	-0.290***	-0.677*	-1.100***
	(0.0806)	(0.303)	(0.105)	(0.0786)	(0.399)	(0.282)
Direction squared		0.251***			0.143	0.240***
		(0.0799)			(0.116)	(0.0740)
Altitude	0.00850	0.00282	-0.00500	0.00885*	-0.00798	0.00361
	(0.00528)	(0.00627)	(0.00600)	(0.00490)	(0.00705)	(0.00570)
Being <2 yrs old, ϕ_0	-1.340**	-1.268**		-1.292**		-1.193**
	(0.600)	(0.569)		(0.621)		(0.557)
Monthly age	0.212**	0.198**		0.211**		0.197**
for children <2 yrs old	(0.0843)	(0.0802)		(0.0821)		(0.0764)
Monthly age squared	-0.0070**	-0.0066**		-0.0070***		-0.0066***
for children <2 yrs old	(0.00282)	(0.00269)		(0.00270)		(0.00254)
Yearly age for children	-0.0507***	-0.0505***	-0.0553***	-0.0469***	-0.0562***	-0.0446***
aged 2 years or above	(0.0121)	(0.0119)	(0.0154)	(0.0136)	(0.0154)	(0.0122)
Age 0 year			-0.644***		-0.652***	
			(0.180)		(0.180)	
Age 1 year			-0.343*		-0.357**	
			(0.178)		(0.174)	
Female	0.0189	0.0105	-0.0533	0.0232	-0.0567	0.0175
	(0.0412)	(0.0400)	(0.0731)	(0.0407)	(0.0721)	(0.0388)
Mother's education	-0.0237**	-0.0181	-0.00517	-0.0234**	-0.00373	-0.0177
	(0.0113)	(0.0111)	(0.0119)	(0.0112)	(0.0117)	(0.0110)
Mother absent	0.0673	0.122	-0.150	0.0767	-0.139	0.137
	(0.137)	(0.138)	(0.127)	(0.134)	(0.128)	(0.135)
Female head	0.00792	0.0430	0.0242	0.00392	0.0363	0.0355
	(0.0726)	(0.0746)	(0.0947)	(0.0701)	(0.0943)	(0.0713)
Household size	-0.0204	-0.0151	-0.0520***	-0.0166	-0.0501***	-0.00940
	(0.0133)	(0.0131)	(0.0158)	(0.0136)	(0.0154)	(0.0121)
Log per capita	-0.0367	-0.0261	-0.0125	-0.0356	-0.00747	-0.0243
household expenditure	(0.0479)	(0.0461)	(0.0748)	(0.0452)	(0.0742)	(0.0435)
Effective window			0.0147**		0.0166**	
			(0.00660)		(0.00645)	
Mills				-0.0942		-0.146
				(0.179)		(0.0990)
Observations	355	355	1,514	1,514	1,514	1,514
R-squared	0.777	0.788				

* $p<0.1$; ** $p<0.05$; *** $p<0.01$. Standard errors clustered at 40 SEAs, the survey's primary sampling unit, in parentheses. Constant and location fixed effects omitted.

Table 4 Estimation results for adults aged 19-69 years.

	(I) OLS	(II) OLS	(III) 1st stage Heckman	(IV) 2nd stage Heckman	(V) 1st stage Heckman	(VI) 2nd stage Heckman
Log distance	-0.957***	-0.974***	0.0901	-0.957***	0.0924	-0.977***
	(0.0631)	(0.0753)	(0.0976)	(0.0614)	(0.0955)	(0.0664)
Direction	-0.502***	-1.586***	-0.219**	-0.502***	-0.276	-1.569***
	(0.0911)	(0.406)	(0.101)	(0.0996)	(0.496)	(0.449)
Direction squared		0.332***			0.0168	0.342***
		(0.114)			(0.141)	(0.122)
Altitude	-0.00215	-0.0105**	-0.000454	-0.00214	-0.000532	-0.0102**
	(0.00525)	(0.00439)	(0.00856)	(0.00510)	(0.00779)	(0.00445)
30-39 years old	0.183	0.184	0.153	0.182*	0.157	0.143
	(0.116)	(0.118)	(0.122)	(0.105)	(0.121)	(0.109)
40-49 years old	0.186	0.223	0.343**	0.185	0.338**	0.132
	(0.138)	(0.134)	(0.148)	(0.138)	(0.148)	(0.160)
50-59 years old	-0.00997	0.0860	0.654***	-0.0118	0.654***	-0.0655
	(0.136)	(0.133)	(0.183)	(0.142)	(0.180)	(0.165)
60-69 years old	0.0972	0.164	0.537***	0.0956	0.534***	0.0306
	(0.139)	(0.138)	(0.198)	(0.140)	(0.197)	(0.171)
19-29 years old × female	0.0370	0.0602	0.401***	0.0359	0.401***	-0.0314
	(0.0898)	(0.0856)	(0.129)	(0.0823)	(0.130)	(0.103)
30-39 years old × female	-0.183	-0.148	0.656***	-0.185	0.649***	-0.290*
	(0.114)	(0.108)	(0.137)	(0.150)	(0.138)	(0.166)
40-49 years old × female	-0.318***	-0.326***	0.352***	-0.319**	0.358***	-0.394***
	(0.114)	(0.118)	(0.134)	(0.130)	(0.134)	(0.122)
50-59 years old × female	-0.000144	-0.0277	0.274	-0.000793	0.266	-0.0845
	(0.134)	(0.126)	(0.216)	(0.141)	(0.215)	(0.135)
60-69 years old × female	-0.397**	-0.367**	0.224	-0.398**	0.226	-0.401**
	(0.170)	(0.156)	(0.248)	(0.173)	(0.249)	(0.177)
Own education	-0.0204**	-0.0163*	-0.0401***	-0.0203*	-0.0398***	-0.00775
	(0.00940)	(0.00867)	(0.0140)	(0.0111)	(0.0139)	(0.00918)
Married	-0.00292	-0.0574	0.568***	-0.00438	0.566***	-0.181
	(0.0676)	(0.0652)	(0.0809)	(0.121)	(0.0779)	(0.130)
Length of residence in	0.00277	0.00336	0.00352	0.00276	0.00355	0.00264
Kabwe	(0.00252)	(0.00206)	(0.00283)	(0.00241)	(0.00282)	(0.00195)
Household size	-0.0105	-0.00998	-0.0688***	-0.0103	-0.0689***	0.00463
	(0.0131)	(0.0127)	(0.0130)	(0.0148)	(0.0129)	(0.0181)
Effective window			0.0161***		0.0171***	
			(0.00609)		(0.00588)	
Mills				-0.0034		-0.284
				(0.196)		(0.210)
Observations	403	403	2,202	2,202	2,202	2,202
R-squared	0.711	0.735				

* p<0.1; ** p<0.05; *** p<0.01. Standard errors clustered at 40 SEAs, the survey's primary sampling unit, in parentheses.

Table 5 The representative BLLs of Kabwe

	All	Male	Female	0-5 years	6-10 years	11-18 years	19-39 years	40-69 years
Mean	15.45	15.77	15.13	23.49	19.80	15.11	10.73	12.14
Percentage of those with								
BLL <5μg/dL	23.2	21.3	25.1	6.0	10.7	21.8	33.7	36.2
5μg/dL≤BLL<45μg/dL	72.0	74.4	69.6	78.4	81.9	76.0	65.6	61.9
45μg/dL≤BLL	4.8	4.3	5.3	15.6	7.4	2.2	0.7	1.9
Observations	4,730	2,320	2,410	768	746	1,014	1,496	706

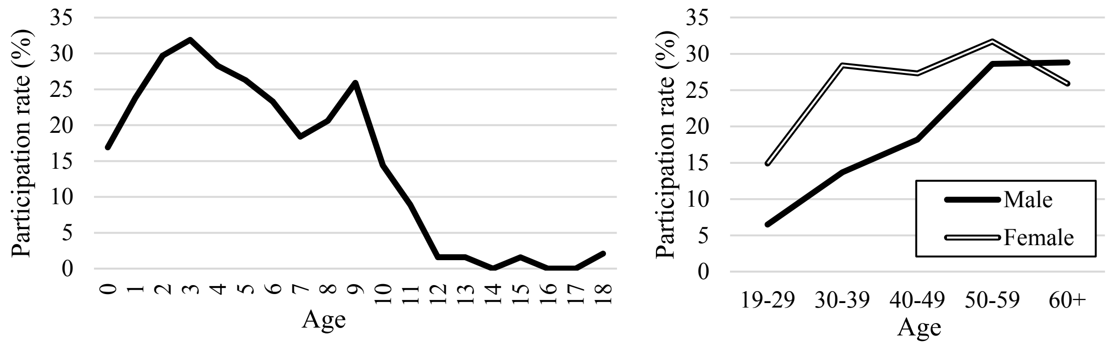

Figure 1 The participation rate by age group (children in left, adults in right)

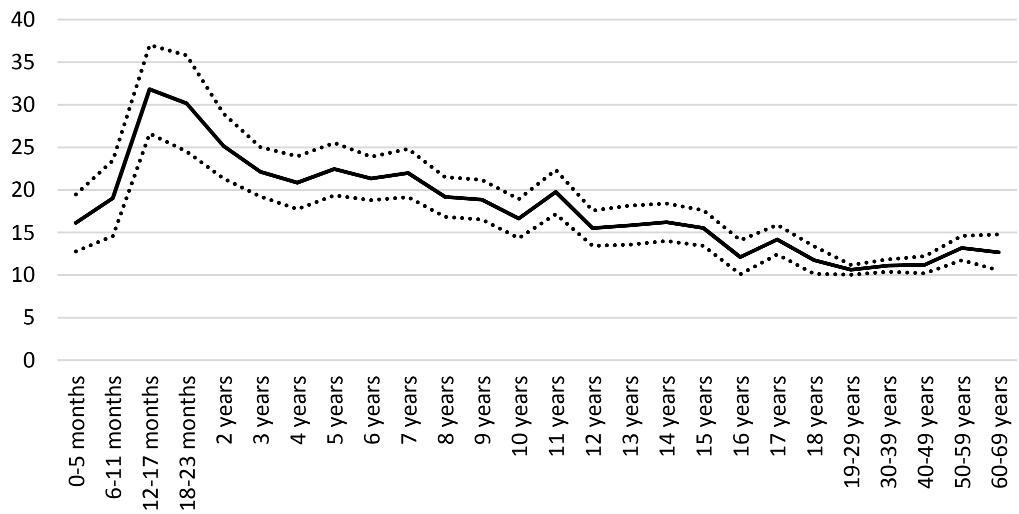

Figure 2 Age-BLL relationship of the predicted BLLs, representative of Kabwe.
Solid line: mean. Dotted lines: 95% confidence interval