Population Association of America Conference
Detroit, April-May 2009

Session 452:
"Heterogeneity and Selection in Infant and Child Mortality"

Controlling for attrition bias and selection at entry in mortality analysis using a two-stage semi-parametric proportional hazard model

Dr Philippe Bocquier[1]

Demography and Population Studies Programme

University of the Witwatersrand, Johannesburg, South Africa

Abstract

All event history analyses make the explicit assumption of independence between censoring and event. Under this hypothesis right censoring due to the time of survey does not create a selection bias. However, when censoring is not independent from the event of interest (e.g. migration in relation to death) then the results suffer from potential bias. This paper presents a way to deal with non-independent censoring, when the same determinants may cause migration and mortality, following the rationale of two-stage regression models controlling for selection biases. The method is tested on longitudinal data collected APHRC Nairobi Urban Health and Demographic Surveillance System (NUHDSS), situated in two large Nairobi slums were circular migration is high. Results confirm the informative censoring hypothesis for out-migration but not for in-migration. The method, though perfectible, produces under-five mortality rates that are consistent with estimates for the poorest quintile in Kenya. Migration is most likely one important strategy adopted by mothers to prevent the risky slum environment to affect their children's health.

[1] Tel.+27 11 717 4355; Fax. +27 11 717 4336, e-mail: philippe.bocquier@gmail.com

1. Introduction

All event history analyses make the explicit assumption of independence between censoring and event. Under this hypothesis right censoring due to the time of survey does not create a selection bias. However, when censoring is not independent from the event of interest, as for migration in relation to death, then the results suffer from potential bias. The independence hypothesis is also broken when left-censoring is selective.

In most demographic data collection systems, migration can rarely be hypothesised to be independent from mortality or any other demographic event for that matter. Migration is a main reason of censoring and should therefore be considered as a major source of disruption in the analysis. Whereas international migration can be neglected at the national level, internal migration is never negligible at the sub-national level. This is especially the case in Demographic Surveillance System (DSS) since they are usually covering a small portion of a country, albeit in great details. The lower the geographical level of the study area, the higher the migration in and out its boundaries.

This paper presents a way to deal with non-independent censoring, when the same determinants may cause out-migration, in-migration and mortality. The method consists in applying to event history analysis (EHA) the rationale of two-stage regression models controlling for selection bias, using instrumental variable (IV) technique. The model is tested on longitudinal under-five mortality data collected in the APHRC Nairobi Urban Health and Demographic Surveillance System (NUHDSS), situated in two large Nairobi slums were circular migration is high (APHRC 2002).

2. Methodological Framework

2.1. Event History Analysis

Most of, if not all, demographic data analysed nowadays contain some form of censoring. The time of survey, or the time of the last round in the case of DSS, is the most common censoring event, and the less troublesome. All event history analysis techniques, from the Kaplan-Meier estimate to the Cox model and its refinement, deal fairly well with this ordinary censoring. Yet other sources of censoring often violate the independence, or non-informative censoring assumption, which states that "*an individual who is censored at* c *should be representative of all those subjects with the same values of the explanatory variable who survive to* c" (Cox and Oakes, 1984).

Informative censoring may be a consequence of informative sampling, i.e. when the sample procedure is such that one part of the sample has a different probability to experience the event as compared to the other. This usually occur when part of the sample is selected on a criteria closely linked to the censoring likelihood. In longitudinal clinical studies, the sample would be biased if observation times are conditioned on certain symptoms or behaviour (Sun & Tong 2009). In this article, we will not deal with informative sampling as it is usually not a major issue in DSS context, where the whole population is followed up.

We will deal here with informative censoring due to some other, endogenous event, i.e. not related to the data collection but rather to the respondents' behaviour. Typically, in DSS context, informative censoring will occur with non-random attrition, e.g. when migration out of the DSS site is dependent on health status acting either as a deterrent or as an incitement to migrate. Informative censoring can also occur exogenously, due to some data collection flow, when some individuals are wrongly excluded from the follow-up by mistake or as a consequence of 'moral hazard in the field', e.g. when field-workers 'out-migrate' respondents to avoid filling forms. We will also control for this bias.

Let (T; C) denote random variables where T is the variable of interest, called the time to event or lifetime, with unknown distribution function (d.f.) F, and C is the random right censoring time with arbitrary d.f. G. In the ordinary survival model framework, T and C are assumed to be mutually independent and, without loss of generality, nonnegative. In the random right censoring model one observes (X; δ) where $X=min(T; C)$ and $\delta=I(T\leq C)$ is the indicator of censoring status:

$$\delta = \begin{cases} 1 & if\ T \leq C \\ 0 & if\ C < T \end{cases}$$

In Cox semi-parametric proportional hazards regression model, we are interested in knowing how the covariates Z affect T:

$$\lambda_{T|Z}(t|\mathrm{z}) = \lambda_0(t)\exp(z\beta),$$

Where $\lambda_0(t)$ is the baseline hazard and the vector β the coefficients which qualify the relationship between T and Z.

In real cases, the function G is often not independent from function F, i.e. it is said to be informative, breaching a common assumption of the survival estimate. Subramanian (2000) showed that the Koziol-Green model assuming:

$$\lambda_{C|Z}(t|z) = p\lambda_{T|Z}(t|z), \text{ with corrective parameter constant } 0 < p < 1$$

often used to deal with informative censoring is restrictive and limits the scope of Cox's model. Yuan (2005) proposed, still in the framework of proportional hazard models, a generalisation of which the Koziol-Green model is a special case:

$$\lambda_{C|Z}(t|z) = \gamma(t,z;\theta)\lambda_{T|Z}(t|z),$$

where the corrective parameter γ is a function of δ, the indicator of censoring status, with logit regression parameters θ:

$$\gamma(t,z;\theta) = exp\{-logit(P(\delta=0|t,z))\} = \exp\left\{-\frac{P(\delta=0|t,z)}{P(\delta=1|t,z)}\right\}$$

Other functions can be applied but the logit function is thought to be the most practical and easily interpretable. In Yuan's model, the entire censoring distribution C is potentially informative. Gather & Pawlitschko (1998) and Zhang & Rao (2004) proposed estimators of the survival function in the case of partially informative censoring, where one function C_0 defines ordinary (non-informative) censoring with d.f. G_0 and another function C_{-1} defines informative censoring with d.f. G_{-1}. The event T is partially observable, since we observe $U=min(T,C_0,C_{-1})$ such that the censoring variable is coded:

$$\delta = \begin{cases} 1 & if\ T \leq C_0 \wedge C_{-1} \\ 0 & if\ C_0 < T \wedge C_{-1} \\ -1 & if\ C_{-1} < T \wedge C_0 \end{cases}$$

Braekers & Veraverbeke (2005) proposed a semi-parametric (Cox) model under partially informative censoring, as a variation of the Koziol-Green model. It is similar to Yuan's model except for the corrective parameter:

$$\gamma(t, z; \theta) = exp\{-logit(P(\delta = -1|t, z))\} = \exp\left\{-\frac{P(\delta = -1|t, z)}{P(\delta = 1|t, z)}\right\}$$

To note, only the informative censoring (*δ=-1*) and the event (*δ=1*), and not the non-informative censoring (*δ=0*), are considered to estimate the corrective parameter *γ*.

Most data sets now include covariates, than may influence both censoring and the event thus invalidating the non-parametric partially informative censoring model. The Koziol-Green model with partially informative censoring would not help either as many data sets now include time-varying covariates in addition to fixed covariates. We would need a variation of the Koziol-Green model such that:

$$\lambda_{C_{-1}|Z(t)}\big(t\big|z(t)\big) = \gamma(t, z(t); \theta)\ \lambda_{T|Z(t)}\big(t\big|z(t)\big),$$

where:

$$\gamma(t, z; \theta) = \exp\left\{-\frac{P(\delta = -1|t, z(t))}{P(\delta = 1|t, z(t))}\right\}$$

Since the covariates *Z* depend on time *t*, the logit becomes a discrete-time logit model with time-varying covariates. This model, to our knowledge, has not yet been developed.

Also, informative right-censoring has attracted most of the attention but informative left-censoring has not been tackled so far. It is however a potentially major source of bias in the case of migration. A selective process is at the origin of most migrations, and this may affect both in- and out-migration flows. Of particular interest to the data analysed in this paper is that health has great potential impacts on the probability to move both ways. Considering the importance of migration flows at the local level, controlling for potential effects of left- and right-censoring is of utmost importance to obtain unbiased rates of demographic events.

2.2. Two-stage equation models to control for selection

Here we adapt to the context of longitudinal data the two-stage equation model that has originally been developed for the control of selection bias in cross-sectional data.

The selection (out-migration) equation takes the form:

$$\lambda_{C_{-1}|Z(t)}\big(t\big|z(t)\big) = \lambda_{C_{-1}0}(t) exp^{z(t)\beta}.$$

The main (mortality) equation takes the form:

$$\lambda_{T|x(t)}(t|x(t)) = \lambda_{T0}(t) exp^{x(t)\beta + \Lambda_{-1}(t)\alpha},$$

where:

$$\Lambda_{-1}(t) = \sum_{j=1}^{N} \lambda_{C_{-1}|Z(t)}(t|z(t)) . I(C_{-1j} \leq t),$$

is the cumulative hazard function computed at the observed informative censoring time C_{-1} only. It is interpreted as a propensity (and not a probability since the cumulative hazard can take value higher than 1) to have out-migrated of the population at risk by censoring time t. The cumulative hazard function is preferred to the inverse of the survival function because of its generalization to renewable event, as is out-migration.

The vectors Z and X are the covariates respectively for the selection and main equations that verify $Z=X+V$. V is a vector of instrumental covariates (variables that can explain the selection but not the event) typically related to data collection issues or to calendar effects that influenced the selection (such as eviction from the population under study).

Taking the log of the main equation and rearranging gives:

$$y(t) = \log\left[\frac{\lambda_{T|x(t)}(t|x(t))}{\lambda_{T0}(t)}\right] = x(t)\beta + \Lambda_{-1}(t)\alpha$$

The equation $y(t)$ is identified if $Z \neq X$, i.e. when the residuals of $y(t)$ are not correlated with instrumental variables $v(t)$ included in vector of covariates $z(t)=x(t)+v(t)$ used to compute the propensity $\Lambda_{-1}(t)$.

Similarly, the propensity to have in-migrated in the population at risk can be introduced to control for potential selection bias due to informative left-censoring. Noting C_{-2} the informative left-censoring time, we have:

$$\Lambda_{-2}(t) = \sum_{j=1}^{N} \lambda_{C_{-2}|Z(t)}(t|z(t)) . I(C_{-2j} \leq t)$$

and:

$$y(t) = x(t)\beta + \Lambda_{-1}(t)\alpha_{-1} + \Lambda_{-2}(t)\alpha_{-2}.$$

Note to the organiser of the session:

This is work in progress. There still is need to:

- *Prove the relation of my model to the partially informative Koziol-Green model. It is possible that my two-stage model and the Koziol-Green model (which is also two-stage, with logit model as first equation) are related, at least in the case of fixed (non-time varying) covariates.*
- *Prove the adequacy of the cumulative hazard function of the informative censoring. Other function could be tested, though it is intuitively the most natural.*
- *Formalise the relation to the two-stage selection model* à la *Heckman used with ordinary (cross-sectional) data. At a later stage, it will probably be necessary to develop our own maximum-likelihood two-stage equation.*

2.3. Migration in longitudinal analysis of mortality

Migration is a major source of population change in DSS. The smaller the study area, the larger the migration is compared to other demographic phenomena. In Nairobi slums, the annual in-migration rate is 27.05% while the out-migration rate is 26.71%. This results in a dramatic turn-over of the slum population. The intensity of the circular migration system just cannot be ignored. The rates are even higher for the children under 5, respectively 36.91% for in-migration and 31.37% for out-migration.

In this situation, migration cannot be considered as mere attrition. Lost to follow-up may occur for other reasons than migration ('moral hazard on the field', as when fieldworkers are 'out-migrating' some people to avoid filling up forms), but this can be only marginal as compared to the large impact of migration. Also, the slum population is constantly renewing itself through in-migration, leaving only a small part of non-movers. It is very likely that a sizable proportion of the migrants are migrating in or out of the slums for health reasons. Actually, even if a small proportion did so, it would have an impact on mortality estimates since most of the population are potential migrants.

Other studies in Africa already showed that sending areas (mainly with prominently rural environment) experience an excess mortality due to people 'returning home to die' (Clark et al. 2007; INDEPTH 2009). It is expected therefore, though it has not been investigated so far, that receiving area experience the opposite: the mortality might be well underestimated due to migrants returning to the sending areas when getting sick. It is not quite sure if this could apply as well to children, considering that it is not they who make the decision to migrate but their parents (most likely their mother). Yet our first hypothesis is that when children get sick, as they often would in a very promiscuous environment susceptible to the spread of infectious diseases, their mother will rather out-migrate to their origin area or send their children to this area for better care. We are not sure if this proves to be an effective way to deal with children diseases, as we do not have follow-up data on the return migrants in their origin area to support that (it might be that the health conditions are not better than in the slums). Our second hypothesis is that the in-migrants parents (mothers) coming with their children would probably not migrate in the slums with sick children, since slums are not suitable place for seeking treatment. Whichever way, our overall hypothesis is that migration is a selective process that leads the healthier to stay in the slums. The mortality should therefore be under-estimated in a context of high circular migration pattern in a poor health environment.

3. Data

Our population of interest is under-five children who lived in the slum settlements at any point in time (as possibly recorded by the NUHDSS) between January 1, 2003 and December 31, 2005, the observation period. For each child, the observation time started at the soonest on January 1, 2003 and at the latest at in-migration or at birth during the observation period. The NUHDSS recorded a total of 5356 births for the period covered by the study out of 16573 children who lived in the two slums before the age of five. The observation time ended at the occurrence of the event of interest (death), at date of censoring due to out-migration or on December 31, 2005. We allowed gaps in the observation time, meaning that children could out-migrate and come back. The number of person years lived by the 16573 children for whom we have complete data was 16125 years, and the outcome of interest was the 253 deaths in the observation period. On average, children lived in the slums for less than one year.

A child is considered to have been born to a recent migrant if born to a mother who migrated to the study community since January 2003.[2] The other mothers are called 'long term resident'. The migration status of the child is also known so that we can cross the two variables (migration status of the mother and of the child) into four categories: born in the slum of long term resident mother, born in the slum of recent migrant mother, not born in the slum of long term resident mother, and not born in the slum of recent migrant mother.

We used demographic and socioeconomic factors known to affect child survival as control variables: those variables directly related to the child (age, sex, ethnicity and place of residence) and those related to the mothers (age, living or deceased, education, and length of stay in the slums). Unfortunately, classic determinants of child mortality such as birth interval and child's rank were not available for all children.

We did not adjust for socio-economic variables because data on these are not routinely collected by the DSS. The available socio-economic data were not collected longitudinally and therefore we could not include them in the model. Attempts to include such variables as time invariant characteristics showed little variation among households. Out of all homes under surveillance, 99.4% have metal roofs (both corrugated and tin/metal sheets), 1.6% have access to their own toilet, 5.1% of the homes to their own source of water (either piped into their houses or the compound), and 7.5% have electricity. Over half (68.2%) of the households have radios while only 6.8% reported to have a television set. Bicycles are found in only 3.4% of the homes while motorcycle ownership is very minimal (0.04%).

4. Results

4.1. Migrations models

The out-migration model (Table 1, last but one column) shows that slum-born have a higher chance to out-migrate the slums than non-slum born. Non slum-born are by definition in-migrants: for them, an out-migration is a second migration at least, whereas for the slum-born, the out-migration is mainly the first. Children born out of the slums whose mother recently migrated have the lower chance to out-migrate. To note, because our model accounts for repeatable events, the same individual may have migrated in and out of the slums.

The gender of the child has no effect on out-migration. Viwandani is the slum where the chance to out-migrate is the highest. The Luhya ethnic group has higher chance to out-migrate than the others. Differences in education of the mother are not significant, except for the missing category, which is obviously not very satisfactory, since it concerns a sizable 19% of the children. Children living in a shelter with tap water have less chance to out-migrate, indicating that the high migration of children in the slums is at least partly motivated by the lack of water if not sanitation.

[2] It should be noted that most mothers who were in the study area when APHRC started running the NUHDSS are also migrants. The analysis would have been most intuitive if we had the variables indicating the date when the people who were there at the beginning of the surveillance migrated into the area so that we could control the analysis for duration of stay in the slum settlement. Unfortunately, the date of arrival in the slum settlement was not collected.

The year effect shows some evidence of an upward trend in out-migration which is also mirrored by an upward in-migration trend. These could be an indication of an increasingly mobile population. However, this could also reflect the improvement of the data collection system, whereby more movements are captured by the year and less people are missed by the data collection team.

Apart from the mother's migration status cum place of birth variable, the notice of demolition is the variable with the strongest effect, with a hazard ratio of 3.219*** [95% CI: 3.042 – 3.435]. Introducing the notice of demolition effect in the mortality equation has no significant effect (results not shown). This confirms that notice of demolition is a good instrument for our two-stage model.

The in-migration model (Table 1, last column) shows that non slum-born of long-term resident mothers have the higher chance to in-migrate. This is a mechanic effect of the children's migrant status: their slum resident mother delivered elsewhere and returned with the baby. This could be indicative either of the lack of (adequate) maternities in the slums or of the strategy of the mothers to deliver out of the slums to minimize the health risks for themselves and their babies.

As for out-migration, gender has no effect. This is indicative that the parents (mothers) do not adopt specific migration strategy according to the sex of their child. Viwandani is not more attractive than Korogocho, although we have seen that more children are out-migrating from Viwandani. This could be an indication that the latter is less suitable for children raising that the former. Viwandani, being situated near the industrial area, is predominantly inhabited by single males and not by families.

The ethnic differences are more or less the same for in- and out-migration but are more pronounced and significant in the case of in-migration. This means first that the circular migration pattern is preeminent among the Luhyas and minorities (grouped as 'other ethnic groups') than for the Kikuyus and Luos, while the Kambas appear to be more stable in the slums. Second, the more significant ethnic differences in the case of in-migration means that the in-migration of the Luhya and other minority groups is increasing as compared to other groups, the Kamba in particular.

Residences with tap water seem to be more attractive, confirming the importance of this amenity in the slums. As for out-migration, sanitation has surprisingly no effect. The negative effect of the mother having a phone is probably due to a selection effect (mothers who can afford a phone will not be attracted by the slums) and not really a deterring effect of this amenity (most phones are actually mobile phone owned by the mother, not land-line attached to the dwelling).

As for out-migration the demolition has an effect on in-migration. This is not a consequence of increased attractiveness of the slums, but rather a data collection issue. Some respondents who moved within the same slum as a consequence of the (notice of) demolition where not always identified as already resident of the same slum. During these years the reconciliation of the identifier of the internal migrants was not effective and a number of these migrants were registered as out-migrants from one place and as in-migrants in another structure, with different identifiers, although they should have retain the same identifiers in the two places of residence. A side effect of this situation is that it inflates the circular migration pattern. Another consequence is that the standard errors in our models are slightly underestimated.

Table 1: In- and out-migration models for under-5 children (2003-2005)

Hazard Ratios	Person-years at risk	Out-migration model	In-migration model
Non slum-born of long-term resident mother [Ref.]	72.60%		
Slum-born of long-term resident mother	18.45%	1.797***	0.695***
		(0.042)	(0.025)
Non slum-born of recent migrant mother	7.91%	0.761***	0.756***
		(0.025)	(0.032)
Slum-born of recent migrant mother	1.03%	2.443***	0.520***
		(0.128)	(0.067)
Female [Ref.]	49.61%		
Male	50.39%	1.000	1.000
		(0.015)	(0.021)
Korogocho [Ref.]	44.38%		
Viwandani	55.62%	1.189***	1.010
		(0.079)	(0.089)
Kamba ethnic group	23.62%	0.960*	0.868***
		(0.021)	(0.027)
Kikuyu ethnic group [Ref]	23.84%		
Luhya ethnic group	17.02%	1.072***	1.071**
		(0.026)	(0.034)
Luo ethnic group	20.53%	1.041*	0.986
		(0.024)	(0.032)
Other ethnic groups	14.99%	1.039	1.093**
		(0.027)	(0.038)
Non educated mother	4.88%	0.956	0.978
		(0.038)	(0.053)
Primary educated mother [Ref.]	57.61%		
Secondary educated mother	18.43%	1.017	0.954*
		(0.020)	(0.027)
Tertiary educated mother	0.29%	1.047	0.994
		(0.122)	(0.152)
Education of the mother missing	18.79%	0.723***	0.983
		(0.018)	(0.033)
Mother alive [Ref.]	99.14%		
Mother dead	0.86%	0.866*	0.807*
		(0.074)	(0.105)
No tap water [Ref.]	42.86%		
Tap water	57.14%	0.782***	1.079**
		(0.017)	(0.034)
No toilet [Ref.]	66.98%		
Own or shared toilet	33.02%	0.975	0.989
		(0.022)	(0.029)
No phone [Ref.]	90.06%		
Phone	9.94%	1.011	0.911**
		(0.027)	(0.034)
Year 2003	35.18%	0.927***	0.876***
		(0.018)	(0.024)
Year 2004 [Ref.]	40.07%		
Year 2005	24.75%	1.165***	1.484***
		(0.023)	(0.041)
No notice of demolition [Ref.]	97.49%		
Notice of demolition (instrumental variable)	2.51%	3.219***	1.537***
		(0.099)	(0.093)
Number of person-years	16125	16125	16125
Number of subjects	16573	16573	16573
Number of events	/	16117	9564

Note: Robust standard errors computed for migration models (repeatable events), which also control for fieldworker effect (results not included in the table). The second column contains the percentage distribution (100%=16125 person-years). Significance level: *=10%; **=5%; ***=1%.

4.2. Mortality models

The results of the mortality model (Table 2, last column) show that the out-migration effect is very strong, with a hazard ratio of 0.115*** [95% CI: 0.052 – 0.255] and shows convexity since the hazard ratio of the squared out-migration propensity is significant at 1.119*** [95% CI: 1.074 – 1.166]. This means that an increase of one unit of out-migration propensity has stronger effect when this propensity is low (e.g. increase from 0 to 1) than when it is already high (e.g. increase from 3 to 4). By contrast none of the in-migration propensity terms are significant.

Therefore, out-migration appears to be selective when measuring the mortality of the children: right-censoring is not independent from death and therefore is a major source of bias in estimating mortality. However, in-migration does not appear to affect our under-five mortality estimates. In other words in-migration does not seem to be selective. This could be an indication that the mothers who are in-migrating in the slums are certainly not motivated by health reasons, whereas the concern for or the situation of their children's health is motivating out-migration.

The direct consequence is that under-five mortality estimate is much higher when controlling for attrition. The under-five crude mortality rate (U5MR) computed using Kaplan-Meier estimate is 60.2 per 1000 and appears to be largely underestimated because of informative censoring. It would be in the vicinity of 157.8 per 1000 when controlling for out-migration propensity only (in-migration being non significant) i.e. two and half time more than the crude estimate. The new, unbiased estimate is closer to what is expected from the living conditions prevailing in the DSA only. U5MR was estimated at 94 per 1000 by DHS-2003 for urban Kenya (though it is probably biased toward middle-class because of under-sampling of urban areas and slums), 117 per 1000 in rural Kenya, and 149 per 1000 in the lowest wealth/assets quintiles in Kenya (i.e. close to our estimate for the slum controlling for attrition). Slums are indeed conducive to higher mortality risks while out-migration (which forms the largest part of attrition) is very likely a response to health hazards encountered in the slums.

Informative (right-)censoring also introduces biases in the measurement of some covariates. The effect of being born in the slums and of being born to recent migrants is largely underestimated when not controlling for migration (Table 2, two last columns). Migration status plays in the same direction for out-migration and for mortality. When taking out-migration into account, mortality risk of the slum-born children almost doubles. This means that the children born in the slums would have on average higher risk to die if they had not out-migrated. This confirms that out-migration is closely linked to the health status of the child.

The excess mortality in Viwandani disappears when controlling for informative censoring, suggesting this was an artifact of out-migration differences between the two slums. The effect of ethnicity remains the same, showing that Luo children have significantly higher mortality risks than other ethnic groups (HR=2.71***). The effect of the death of the mother, which was considered only if it occurred during the observation time of the child, becomes significant (HR=2.39***) after controlling for attrition. Despite its effect on migration, access to tap water has not effect on mortality, nor do access to sanitation or phone. The effect of the year is not significant showing no change in the mortality level over the admittedly short period of observation.

Table 2: Mortality models for under-5 children (2003-2005)

Hazard Ratios	% person-years at risk, or mean value	Mortality model without controlling for migration	Mortality model controlling for migration
Non slum-born of long-term resident mother [Ref.]	72.60%		
Slum-born of long-term resident mother	18.45%	1.860***	3.036***
		(0.312)	(0.878)
Non slum-born of recent migrant mother	7.91%	0.576	0.381**
		(0.236)	(0.177)
Slum-born of recent migrant mother	1.03%	3.338***	5.824***
		(0.963)	(2.518)
Female [Ref.]	49.61%		
Male	50.39%	1.184	1.177
		(0.146)	(0.153)
Korogocho [Ref.]	44.38%		
Viwandani	55.62%	0.733**	0.890
		(0.107)	(0.153)
Kamba ethnic group	23.62%	0.712	0.667
		(0.158)	(0.180)
Kikuyu ethnic group [Ref]	23.84%		
Luhya ethnic group	17.02%	1.352	1.477*
		(0.299)	(0.349)
Luo ethnic group	20.53%	2.627***	2.714***
		(0.498)	(0.519)
Other ethnic groups	14.99%	0.689	0.713
		(0.211)	(0.216)
Non educated mother	4.88%	1.083	1.043
		(0.405)	(0.420)
Primary educated mother [Ref.]	57.61%		
Secondary educated mother	18.43%	1.102	1.106
		(0.173)	(0.205)
Tertiary educated mother	0.29%	4.178	4.710
		(41.234)	(46.011)
Education of the mother missing	18.79%	0.850	0.634*
		(0.225)	(0.175)
Mother alive [Ref.]	99.14%		
Mother dead	0.86%	2.836	2.392**
		(13.198)	(1.012)
No tap water [Ref.]	42.86%		
Tap water	57.14%	1.125	0.877
		(0.236)	(0.178)
No traditional toilet [Ref.]	66.98%		
Traditional toilet	33.02%	1.169	1.164
		(0.208)	(0.208)
No phone [Ref.]	90.06%		
Phone	9.94%	1.143	1.147
		(0.231)	(0.220)
Year 2003	35.18%	1.294*	1.058
		(0.194)	(0.173)
Year 2004 [Ref.]	40.07%		
Year 2005	24.75%	0.977	1.333
		(0.156)	(0.337)
Out-migration propensity	2.109		0.115***
			(0.047)
Out-migration propensity squared	7.487		1.119***
			(0.023)
In-migration propensity	1.535		1.453
			(0.768)
In-migration propensity squared	3.217		0.904
			(0.088)
Number of person-years	16125	16125	16125
Number of subjects	16573	16573	16573
Number of events	/	253	253

Note: Bootstrap standard errors (200 replications). The second column contains the percentage distribution (100%=16125 person-years) or the mean value of migration propensity scores. Significance level: *=10%; **=5%; ***=1%.

5. Conclusion

The results presented in this paper clearly show that, conforming to our hypothesis, censoring cannot be considered to be independent from the event of interest, death in our case. Out-migration is a major source of disruption at low geographical level, though in-migration does not seem to bias under-five mortality estimates. Therefore, migration should be given utmost attention in demographic surveillance systems.

Although the method proposed in this paper needs to be refined and its results compared with other methods proposed to control for selection biases, the paper demonstrated the practicability of the proposed procedure to measure demographic events controlling for migration. Using this procedure, the estimations reflect better the conditions prevailing in the study area. The method produces under-five mortality rates that are consistent with the national estimates for the poorer quintile and allows non-biased identification of the determinants of mortality. The results also confirm our first hypothesis that parents (or mothers) will send their children out of the slums when they get sick or are susceptible to diseases. However, contrary to our second hypothesis, the in-migrant children do not prove to be selected upon their good health. In other words, slum in-migration does not appear to follow any health-related strategy or behavior, whereas slum out-migration does.

The results call for further analysis of the determinants of both migration and morbidity in Nairobi slums, controlling for selection by migration. We would want to have a more precise measure of prevalence of diseases among children followed in the demographic surveillance system and to know better about the response of the parents (mothers) to the diseases of their children, in relation to migration, through qualitative and quantitative surveys.

Acknowledgements:

I would like to thank Eliya Zulu and Alex Ezeh of APHRC for providing me with the data and for the necessary technical support. My discussion with Adama Konseiga and Woubalem Zewdu motivated an earlier version of this work. This work would not have been possible without the Wellcome Trust which gave generous support to the Urbanisation, Poverty and Health Dynamics in Sub-Saharan Africa (UPHD) programme.

6. References

APHRC. (2002b) Population and health dynamics in Nairobi Informal Settlements. African Population and Health Research Center, Nairobi Kenya.

Braekers, R. & Veraverbeke, N. (2005). "Cox's Regression Model Under Partially Informative Censoring". *Communications in Statistics – Theory and Methods*, 34: pp1793-1811.

Clark, S.J., Collinson, M.A., Kahn, K., Drullinger, K., Tollman, S.M. (2007). "Returning home to die: Circular labour migration and mortality in South Africa." *Scandinavian Journal of Public Health* 35(supplement 69): pp35-44.

Collinson, M.A., Adazu, K., White, M., Findley, S.E. (2009). Households Unfolding: INDEPTH Network Perspectives on the Dynamics of Migration, Health and Livelihoods. INDEPTH. To be published.

Cox, D.R. & Oakes, D. (1984). Analysis of Survival Data. London: Chapman andHall.

Gather, U. & Pawlitschko, J. (1998). "Estimating the survival function under a generalized Koziol-Green model with partially informative censoring. *Metrika*, 48: pp189-207.

Subramanian, S. (2000). "Efficient estimation of regression coefficients and baseline hazard under proportionality of conditional hazard". *Statistical Planning and Inference*, 84(1 2): pp81-94.

Sun, L. & Tong, X. (2009). "Analyzing longitudinal data with informative observation times under biased sampling". *Statistics and Probability Letters* (in press).

Yuan, M. (2005). "Semiparametric Censorship Model with Covariates". *Sociedad de Estadistica e Investigacion Operativa Test*, 14(2): pp489-514.

Zhang, H & Rao, M.B. (2004). "On generalized maximum likelihood estimation in the proportional hazards model with partially informative censoring". *Metrika*, 59: pp125-136.