Multilevel multiple imputation of missing birth weights in developing countries: Analysing neonatal and post-neonatal mortality

Andrew R. Channon*

Introduction

Birth weight is acknowledged to be one of the most important indicators of health in both childhood and later life (McCormick, 1985; Barker, 1992). However, in many developing countries a large proportion of infants are not weighed at birth and those that are weighed are not representative of all infants (Miller *et al.*, 1993). Therefore, the use of birth weight when modelling health outcomes excludes a significant proportion of infants and leads to biased results.

There are limited options available to the researcher if birth weight is to be used in analyses in developing countries. A simple option is to use a proxy variable for birth weight. One such proxy variable that has been used by some authors is the mother's perception of the size of her baby at the time of birth (Rodrigues and da Costa Leite, 1999; Magadi *et al.*, 2007). The size of the baby is classified into five categories, ranging from very small to very large and is routinely collected in Demographic and Health Surveys (DHS). However, a consequence of using mothers' perceptions of size as a proxy in regression models is a loss of the rich continuous birth weight information, being replaced by only five categories. It is preferable to retain the continuous birth weight information if at all possible.

The main aim of this paper is to assess the results of using multiple imputation to impute missing birth weights in developing countries. Due to the hierarchical structure of many surveys (including the DHS) the imputation procedure needs to account for the survey design in order to provide reliable results. Traditional multiple imputation procedures do not account for this hierarchical structure. However, a

* Southampton Statistical Sciences Research Institute and Division of Social Statistics, University of Southampton, UK

method developed by Carpenter and Goldstein (2004) accounts for the complex survey design and imputes continuous measures, but this method has not been applied to non-simulated missing data in surveys where the proportion of missing information may be very high. This study uses this method to impute birth weight in three countries and these imputed datasets are then used to analyse the determinants of neonatal and post-neonatal mortality.

Data and Methods

DHS data are used from Cambodia (2000), Malawi (2000) and Kazakhstan (1999). These countries were chosen due to the different proportions of infants without a recorded birth weight in the survey. In Kazakhstan only 2.9% of infants did not have a birth weight recorded in the survey, while the corresponding proportions in Malawi and Cambodia are 55.9% and 84.1% respectively. Infants born in the five years before the date of the survey had detailed information collected, including the perception of size and their birth weight, if available.

There were two stages in this analysis:

1. To assess if the multilevel multiple imputation method is a viable method to use for birth weight, firstly only those infants with a recorded birth weight were considered. Different proportions (between 5% and 25%) of these infants had their birth weights recoded as 'missing'. The multilevel multiple imputation procedure was applied. To assess the method:
    - The imputed birth weights were compared to the actual birth weights and the average difference in birth weight was calculated.
    - A logistic regression model of both neonatal and post-neonatal mortality using the survey recorded birth weights was compared to the logistic models using the imputed data sets. The difference in parameter values was calculated.
2. The multilevel multiple imputation method was applied to the full dataset (all infants, with and without a recorded birth weight) and birth weights were imputed for those infants with missing birth weights. Logistic regression was conducted to analyse the determinants of neonatal and post-neonatal mortality

in the three countries, and the results were compared to models using only those infants with a recorded birth weight.

The missing birth weights were imputed using the technique developed by Carpenter and Goldstein (2004), termed multilevel multiple imputation. A macro has been developed which implements this technique in the statistical software package, MLwiN (Institute of Education, 2005). The macro fits a multilevel multivariate imputation model using Bayesian methods with uninformative priors, using MCMC methods. Variables used in the imputation model include gender, survival status, educational level of the parents and mother's perception of size. Perception of size is seen to be highly related to birth weight and is available for most infants in the survey, and thus is a good variable to use in the imputation model.

Preliminary Results

A brief summary of the key results currently obtained indicate that:

Results of Simulation

- The average values of the imputed birth weights across all imputations is very close to the true value of the reported birth weight, even with up to 25% of the birth weights set to missing.
- Accuracy of the imputations does not decrease as the proportion of missing data increases.
- Models of neonatal and post-neonatal mortality using the imputed birth weights produce parameter estimates which are very close to the results obtained when using the actual birth weight. The similarity of results declines as the proportion of missing data increases, although with 25% missing data the results are still close.

Results of Mortality Analysis

- Imputing the birth weights in each of the countries and modelling neonatal and post-neonatal mortality produces different parameter estimates compared to those obtained when using only the infants with a recorded birth weight. This

is the case even in Kazakhstan where there are only 2.9% of infants without a recorded birth weight.

- The parameter estimates of birth weight and other covariates in the model are closer to those which are expected compared to the complete case analysis. There is a strong positive relationship between birth weight and mortality and this relationship is stronger after imputation.

Conclusions

Multilevel multiple imputation is a good method to impute birth weights in countries where there are some infants who are never weighed at birth. The parameter estimates obtained during modelling are closer to those expected than when only using those infants with a recorded birth weight. This method allows continuous birth weight to be imputed, thereby allowing a large range of outcomes to be investigated, especially when birth weight is needed as a control variable in a model. Overall, this method allows the better and more accurate analysis of child health outcomes in developing countries where there is missing information.

References

Barker, D., Ed. (1992). Fetal and infant origins of adult disease. London, BMJ.

Carpenter, J. and H. Goldstein (2004). Multiple imputation in MLwiN. www.missingdata.org.uk.

Institute of Education (2005). MLwiN Version 2.02.

Magadi, M., A. Agwanda, F. Obare and N. Taffa (2007). "Size of newborn and caesarean section deliveries in teenagers in Sub-Saharan Africa: evidence from DHS." Journal of Biosocial Science **39**(3): 175-187.

McCormick, M. (1985). "The contribution of low birth weight to infant mortality and childhood morbidity." The New England Journal of Medicine **312**(2): 82-90.

Miller, J. E., N. Goldman and L. Moreno (1993). "An evaluation of survey data on birthweight and prematurity status." Social Biology **40**(2): 131-146.

Rodrigues, R. and I. da Costa Leite (1999). Determinants of neonatal mortality in Brazil. Population Association of America. New York.