# Integrating area-based and national samples in birth cohort studies: the case of Life Study.

by

Harvey Goldstein, Francesco Sera[1], Peter Elias[2] and Carol Dezateux

Life Course Epidemiology and Biostatistics

Population Policy and Practice Programme

UCL Institute of Child Health, 30 Guilford Street London WC1N 1EH

## Abstract

The most recent UK birth cohort study, known as 'Life Study' was a longitudinal study planned to involve some 80,000 babies and comprised two components. The largest, the 'Pregnancy Component' was to consist of around 60,000 pregnant women who were to be recruited when attending for a routine antenatal ultrasound at selected maternity units in England. The other component, the 'Birth Component' was to be a random sample of intended size 20,000 live births across the UK. Recruitment to the cohort was to take place over a period of four years starting in 2015. Innovative sampling procedures had been designed and tested and a synthetic dataset produced with similar characteristics to the anticipated survey data was produced to study the performance of the sampling procedures and explore analysis strategies.

[1] Present address:

Research Fellow in Medical Statistics

Department of Social and Environmental Health Research,

London School of Hygiene and Tropical Medicine

London, WC1, UK

[2] Professor

Warwick Institute for Employment Research

University of Warwick, Coventry, CV4 7AL, United Kingdom

Tel: +44 (0)24 7652 3284

This research note describes the proposed sample design, and discusses how the two components were to be integrated to provide a consistent dataset for users. Approaches to the provision of suitable sampling weights and modelling approaches are also presented. Lessons are drawn for designs of future cohort studies.

## Keywords

Cohort studies, Longitudinal studies, Research design, Probability Sampling, weighting, attrition, non-response

Address for correspondence:
h.goldstein@bristol.ac.uk

## Introduction

The most recent UK birth cohort study, Life Study, began in 2011 as a longitudinal study of pregnant women and births in the UK with recruitment designed to take place over a four year period. It comprised two components, each representing different sampling and recruitment strategies, data from each of which was to be integrated into a single dataset for interdisciplinary research investigating causal mechanisms while enabling generalisability to the UK population. It was funded principally by the Economic and Social Research Council (ESRC). In July 2015, six months after recruitment had started in the first maternity centre, the ESRC council announced its decision to withdraw funding from October 2016, which was within two weeks of the planned opening of the second centre. (Dezateux et al., 2016a). There clearly is interest in understanding why it was decided to close down Life Study, beyond the brief formal statement issued by the funders, and the reader is referred to the summary report from the Life Study Scientific Steering Committee (Dezateux et al 2016a) for further information. Further discussion of this is not relevant to the purpose of the present research note which focusses solely on the study design and lessons that may be learnt from this for the future.

This sampling design was substantially the one recommended in two reports to the principal funder (Bynner et al., 2007, 2009), as set out in the protocol approved by the principal funder following a competitive process, international peer review and assessment by an independent international scientific panel (Dezateux, 2016b). The sampling design was subsequently discussed and refined by a methodology advisory group set up by and reporting to the Life Study Scientific Steering Committee.

One of the components - the 'Birth Component' - was planned as a sample of approximately 20,000 births selected at random, uniformly over the period of recruitment, with the sample being selected from the UK birth register. This component had similar intent to previous UK birth cohort studies - such as those starting in 1946, 1958, 1970 and 2001, namely to recruit a large and representative sample of births from across the UK.

The second component - the 'Pregnancy Component' - was planned as a sample of approximately 60,000 pregnant women approached when attending for routine antenatal ultrasound at maternity units in three geographically defined areas of England. A detailed description of the protocols for both components, including power calculations and sampling strategies, is given elsewhere (Dezateux et al., 2016b). Full documentation of the study questionnaires and materials can be found at https://www.lifestudy.ac.uk/resources, and these will be useful for readers who wish to gain a deeper understanding of the overall rationale and content of the study.

In this research note our concern is to explain the specific rationale for the formal sampling design of the study and to illustrate, using a synthetic dataset, different approaches to handling the integrated dataset by analysts, including the production of weights, the use of models that explicitly incorporate design factors, and procedures for handling missing data.

This two component design of Life Study was motivated by the need to provide information that was both representative of a population in time, for example to provide comparisons with previous birth cohorts sampled from the whole of the UK, as well as to enable scientific information to be collected before birth using innovative measures and approaches in order to provide insight into causal mechanisms underlying child development and health associated with pregnancy characteristics. Both components were conceived of as part of a single study with the aim of creating, as far as was practicable, a harmonised dataset and follow-up occasions. The pregnancy component is important since the biological and other data required are generally only possible to obtain through working in close collaboration with maternity units (as there is no sampling frame for pregnancies) and by setting up dedicated centres in which a wide variety of detailed measurements and observations can be obtained which would not otherwise be possible in a home setting. Previous pregnancy cohorts that have recruited mothers in pregnancy, have successfully used a similar approach, including, in the UK, the Avon Longitudinal Study of Parents and Children (ALSPAC) (Boyd et al., 2013) and the Born in Bradford study (Wright et al., 2013). In Goldstein et al, (2015) there is a detailed discussion around the role of the representativeness of samples and how non-probability based samples can be used to enhance scientific causal modelling objectives when combined with population based data.

Life Study was designed to be interdisciplinary from the outset. Hence it was considered desirable to make data from both components available for analysis as a single integrated dataset as this would provide additional precision for both causal analyses and population estimates. A major objective of Life Study was to enable analyses of causal mechanisms underpinning relationships between a range of exposures and later child outcomes, mostly derived from the larger, more intensively phenotyped, participants in the pregnancy component. The selection of maternity units was informed by the modelling of routine data on births linked to demographic information from the 2011 UK Census. This was undertaken by the Small Area Health Statistics Unit at Imperial College London. The initial criteria used to select potential units included measures of ethnic and social diversity, scale as well as geographical spread of mothers comprising the antenatal population served by each maternity unit.

The remainder of the paper is organised as follows. In the next section, we will briefly refer to the creation of a synthetic dataset created to test different procedures for data modelling and presentation. Following that there is a section that describes how weights to handle the differential sampling across components and also non-response, can be defined and calculated. The final section describes some analyses, using the synthetic dataset, to illustrate different approaches to statistical modelling.

## Developing a synthetic dataset

Given the premature closure of the Study, actual study data are not available and hence we report the use of a synthetic dataset produced in order to develop and test sampling and analysis methodology. Since Life Study itself did not proceed beyond the initial pilot stage, we use this synthetic dataset to illustrate the design and analysis issues set out above. We do not address the issue of sampling costs, although in practice this will be important. Birth statistics for England and Wales are produced by the Office for National Statistics (ONS, 2013) and those published for the 2012 calendar year were used to generate a synthetic dataset, based upon marginal and pairwise table distributions. Table 1 lists the variables used.

| **Table 1. Synthetic dataset variable definitions** |
| --- |
| Sex |
| Area of usual residence of the mother (Local Authority District (LAD)) * |
| Area of usual residence of the mother (Electoral Wards) ** |
| Type of birth registration (Within Marriage, Joint Registrations same address, Joint Registrations different address, Sole registrations) |
| Age of the mother at birth (years) (Under 20, 20-24, 25-29, 30-34, 35-39, 40-44, 45 and over) |
| Age of the father at birth (years) (Under 20, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50 and over) |
| National Statistics Socio-economic Classification (NS-SEC) (1.1, 1.2, 2, 3, 4, 5, 6, 7, 8) *** |
| Ethnicity of baby (White, Black, Asian, Others) **** |
| Number of previous live-born children (0, 1, 2, 3+) |
| Multiple births (Singletons, Twins, Triplets) |
| Birth weight (Under 1,500 g, 1,500-1,999, 2,000-2,499, 2,500-2,999, 3,000-3,499, 3,500-3,999, 4,000-4,499, 4,500-4,999, 5,000 and over) |
| Month of birth (January to December) |
| * Local authority districts (LAD) is a generic term to describe the 'district' level of local government in the United Kingdom. It includes non-metropolitan districts, metropolitan districts, unitary authorities and London boroughs in England; Welsh unitary authorities; Scottish council areas; and Northern Irish district council areas. We considered 346 LAD in England and Wales. |
| **Electoral wards/divisions are the key building blocks of UK administrative geography. They are the spatial units used to elect local government councillors in metropolitan and non-metropolitan districts, unitary authorities and the London boroughs in England; unitary authorities in Wales. We considered 8565 Electoral wards in England and Wales with 2012 boundaries. |
| ***1 Higher managerial and professional occupations; 1.1 Large employers and higher managerial occupations; 1.2 Higher professional occupations; 2 Lower managerial and professional occupations; 3 Intermediate occupations; 4 Small employers and own-account workers; 5 Lower supervisory and technical occupations; 6 Semi-routine occupations; 7 Routine occupations; 8 Never worked and long-term unemployed. This uses the combined method based on the most advantaged NS-SEC of either parent rather than just the father's socio-economic classification. |
| **** White (White British, White Irish and any other ethnic background); Black (African, Caribbean and any other Black background); Asian (Bangladeshi, Indian, Pakistan and any other Asian background); Other (All Mixed groups, Chinese, any other ethnic group). Based on Census categories. |

The initial synthetic dataset comprised 2,918,696 records (live births), being the estimated number that would accrue over a four year period, in England and Wales, based on ONS Birth Statistics.

The pregnancy component sample was built as follows. A multivariable logistic model was used to predict the probability of being included in the pregnancy component; as predictors we used four of the variables listed in table 1, namely: ethnicity, SES, maternal age and birth weight. We fixed the coefficients in order to include higher proportions of ethnic minorities, low SES, low maternal age and low birth weight than in the population. The intercept and the coefficients in the reference category of each of the four predictors were set to have a pregnancy component sample equal to 121,802, that is four years live births, which is 4.2% of the total synthetic dataset. As discussed below, with a non-response rate equal to 50% this will result in approximately 60,000 live births, which was the planned sample size of the pregnancy component.

Under the assumption of non-overlap between the two components the target population of the birth component was composed of the remaining 2,796,894 = (2,918,696-121,802) four years' worth of estimated live births in the synthetic dataset. The birth component sample was randomly selected without replacement, to have a sample size equal to 31,941 (sampling fraction equal to 0.0114=31,941 /2,796,894). This sample size was set in order to have an observed sample size of 16,000 for the birth component in England and Wales, with the remainder of the birth component to be sampled from Scotland and Northern Ireland. The total sample size (Pregnancy +Birth, before non-response) was thus equal to 153,743.

Based on the experience of similar large scale cohort studies, an initial conservative estimate of a 50% response rate was assumed for Life Study and so this is assumed for the synthetic dataset, with an oversampling of ethnic minorities as proposed in the Pregnancy component outlined above. After allowing for 50% (randomly occurring) non-response, the final synthetic dataset consisted of approximately 60,000 in the pregnancy and 16,000 in the birth component.

For the pregnancy component we created the equivalent of (post stratified) design weights based upon the distributions in the full dataset. For the purpose of estimating weights any

clustering within the pregnancy component is ignored. In this dataset we have a table consisting of 256 cells generated by the four variables, Age of the mother, National Statistics Socio-economic Classification, infant ethnicity, and Birth weight. The weight for each cell *j* is set to

$w_j = (\frac{N_j}{N})/(\frac{n_j}{n})$

where upper case denotes the generated population and lower case the sampled records. Note that the post stratified design weights for the pregnancy component are proportional to $\frac{121{,}802}{2{,}918{,}696} = 0.042$ which reflects the fact that, treated as a stratum of the population it includes only 4.2% of live births. For the birth component the design weights are simply the inverse of the relevant sampling fraction. The weights were scaled to add to the total sample size. Thus, overall the sampling weights have mean of 1 with a computed standard deviation equal to 1.85. This would imply a design effect of 4.42. The term 'design effect' (Kish, 2014) is used in the standard way to mean the ratio of the variance (of the estimate) associated with the actual sample to the variance associated with a simple random sample of the same size.

## Non response at the first sweep

The first sweep was planned to be the baseline examination at 28 weeks' gestation for the pregnancy component and at 6 months after birth for the birth component.

To incorporate possible bias within the synthetic dataset due to non-response, an 'informative' non-response mechanism depending on the available 'auxiliary' data was simulated. A main effects logistic model for the propensity to respond assumed differential responses for the categories of the four auxiliary variables. In particular, coefficients associated with each variable were fixed to have higher proportions of non-response within ethnic minorities, low SES, low maternal age and low birth

For each of the two components, the final weights have been calculated as the product of sampling (or post stratified design) weights divided by the probability of response, and then scaled to the obtained sample size of 77,202, which differs slightly from the intended size of 76,000.

The resulting weights represent the theoretical weights based upon the known expected values derived from the sampling and non-response mechanism. We can also compute

weights based upon the achieved sample characteristics (after non response adjustment) and the known population values. This can be done directly for each cell of the multiway table (amalgamating very small cells) or via the logistic prediction models described above. The leads to a set of weights with mean 1 and standard deviation 2.03.

## The use of weighting and covariate adjustment: efficiency and causal modelling

The data for Life Study essentially were to arise from five basic strata, namely the four constituent countries of the UK (birth component) together with the pregnancy component which was to be drawn from maternity units based in three English National Health Service Trusts. Within each stratum the design was intended to produce a sample where each member had the same selection probability, together with weights that would reflect the proportion of the whole UK population of births together with non-response adjustment, as described above. The design of Life Study required some oversampling for Scotland, Wales and Northern Ireland so that, should separate estimates be required for these countries, they would have an acceptably small standard error. The proposed distribution for the birth sample was as follows: England n=15500, Wales n=1500, Scotland n=1500, Northern Ireland n=1500. This would have resulted in a relative oversampling ratio for Wales, Scotland and Northern Ireland compared to England of 2.0, 1.2 and 2.4 respectively.

Standard data analysis packages will carry out weighted analyses and for many purposes this will be adequate where inference is required for the (national) population. The extent of clustering in the design will allow the calculation of a design effect for the birth component which can be used to adjust standard errors. An alternative is to fit a 2-level model where the clusters are level 2 units, and this is generally to be preferred.

As described above, the design weights are adjusted for differential non-response using post-stratification based upon known population birth characteristics. We discuss the case of missing data and attrition below, but ignoring such possibilities for now, for the purpose of providing descriptive population estimates, the use of these weights is straightforward. We note that the addition of the pregnancy component to the birth component, *for the purpose of providing population estimates*, is estimated to add only between 5% and 10% efficiency, since the pregnancy component provides just 4% of the population of births,

which is reflected in the low overall set of weights for this component. Thus, for the design effect of 4.42 quoted above, the effective sample size for estimating the mean, becomes 17,750 (16,841 from the births component + 949 from the pregnancy component). As we have already suggested, such population estimates will be useful, for example, when making comparisons with previous birth cohort studies both in the UK and elsewhere.

When fitting statistical models for the purpose of making causal inferences, we could also use these weights although, as we discuss below, it is both more flexible and more efficient if a full statistical modelling approach is used that treats the post stratification variables, used to derive the weights, as auxiliary variables or covariates within the statistical model. This allows us to incorporate all the information in the pregnancy component data without down-weighting it, so giving an efficient analysis. To illustrate the properties of the two approaches – weighting, or adjusting by strata defined by auxiliary variables, - we consider a scenario in which the researcher is interested in analysing a linear relationship between a continuous exposure *x,* for example the duration of breastfeeding in the first six months following birth, and a continuous outcome *y,* for example weight adjusted for height at twelve months. The linear relationship has a coefficient equal to 0.20 within each stratum defined by the design auxiliary variables. We thus have the model

$$y = \alpha + 0.20x + f(z) \qquad (1)$$

where $z$ represents the set of auxiliary variables that not only adjust for the design, but also contain variables such as ethnicity and birthweight that are relevant confounders. For simplicity we assume a main effects model without interactions.

Table 2 shows the results of fitting the model fitting auxiliary variables with and without using weights, for the model given by (1). Note that in deriving these weights we have ignored the post-stratification information provided by the auxiliary variables, and so the results presented give a 'worst case' scenario.

Given the fact of the much larger sample size of the pregnancy component, in analyses of simple regression models, while the parameter estimates themselves are little changed, in the present case and also more generally, it leads to an increase in efficiency. As shown in table 2 the standard error of the coefficient estimated in the model adjusted for strata design is 0.0024 in the combined dataset and 0.0053 in the birth component alone. Using the weights, since the pregnancy component represents only a small component of the population, the estimated standard error shows that this reduced the efficiency by a factor

of 4.5. This illustrates the advantage of a full modelling approach since it allows adjustment for the auxiliary variables relevant to data analyses that explore scientific hypotheses, for example those relevant to causal pathways. Interactions in such models between designated causal variables and the auxiliary variables will also often be of interest, for example in studying whether the causal relationships vary by ethnic group, mother's age, etc. As here, it is often the case when auxiliary design variables are used, that these are indeed also relevant for inclusion in the analyst's models of interest. Where one or more such variables are not of interest, but are associated with the response variable(s), and the analyst wishes to have model estimates that average or 'marginalise' over these variables, this is perhaps most readily achieved by omitting those variables from the model and introducing corresponding weights to compensate.

**Table 2. Coefficients and standard errors estimated in a linear regression (model 1) adjusted by strata design and auxiliary variables and using weights**

| | Adjusted by strata/auxiliary variables | | Using weights (worst case scenario) | |
|---|---|---|---|---|
| | Coefficient | S.E. | Coefficient | S.E. |
| Birth component | 0.206 | 0.0053 | 0.206 | 0.0053 |
| Combined | 0.201 | 0.0024 | 0.205 | 0.0050 |

For a discussion of the relative merits and uses of population representative samples see Goldstein et al. (2015).

## Models for attrition and item non-response

It is anticipated that there will be some loss of data from the study due to item non-response in the questionnaires, failure to obtain measurements, or to cohort study members not attending subsequent sweeps (attrition). A detailed discussion of how to deal with this is given by Goldstein (2009), but briefly is as follows. As described above, the synthetic dataset incorporates missing data values due to non-response where the failure to respond is a function of the stratification variables and other variables that, in the case of attrition, were available at the first wave of data collection. Although there was not time

fully to explore ways of dealing with such missing data, we will describe the overall approach that was intended.

In practice, as a study progresses, those variables associated with the propensity to respond would be identified and then used to adjust for response biases. In the case where a complete record for an individual is missing, for example due to subsequent attrition, a data analyst may wish to use procedures, such as propensity score matching or inverse probability weighting (see for example, Pearl, 2009), or carry out a full imputation modelling, conditioning on the variables associated with propensity to respond. Thus, if an item in a record is missing, a form of multiple imputation is generally recommended so that maximum efficiency can be maintained and bias minimised. Auxiliary variables associated with missingness can be incorporated and missing items and attrition can be handled simultaneously. When using multiple imputation, users will obtain several, typically at least ten, completed multiply imputed datasets. The model of interest is then fitted to each one and parameter estimates combined according to simple rules (Carpenter and Kenward, 2013). While it is possible to carry out the imputation prior to releasing data for secondary analysis, this does have certain drawbacks, especially since it would be impossible to anticipate all the analyses that users might wish to do and to include all the relevant variables in the imputation. The imputed datasets would therefore not properly reflect the variables in such analyses and this is known to create problems due to a lack of 'congeniality' (Carpenter and Kenward, 2013). Instead, therefore, Life Study was proposing that suitable materials be provided for users to enable them to carry out their own efficient, and theoretically sound, imputation-based analyses. One of these was to have been based upon the proposal by Goldstein et al (2014), who present a Bayesian procedure to carry out such analyses. Appropriate software could be made available to potential users of combined survey datasets as part of an 'access and analysis' package.

If the data are being used to make population estimates then the data records will also have weights attached and these would need to be incorporated. Carpenter and Kenward (2013) discuss how this may be done and further work on this is currently being carried out (Goldstein, Carpenter and Kenward, 2016).

## Response rates

A persistent feature, observed internationally, is the decline in response to population surveys. We have discussed how this can be tackled by carrying out adjustments based upon nationally available data, although when the response rate becomes very small this may not become practical. We can often also do this using administrative data for small areas or institutions such as those defined by women attending a set of maternity units. This is likely to be an important future consideration and thus should be a feature of any study design. For components of any study that are intended to represent real populations, there will typically be administrative data that comprehensively cover the population of interest. In the case of Life Study, as described, comprehensive data on live births were available. For components that sample from institutions obtaining such institution-level data may often be problematical, and for such studies ensuring at the planning stage that these are made available is important. This may require, for example in the case of women attending a maternity unit, ethical approval as well as suitable record systems being available.

Acquiring such auxiliary data, while important, does not imply that obtaining high response rates and minimising attrition are unimportant. Concentrating data collection within institutions such as maternity units or schools may often have advantages in promoting high response rates by increasing participant motivation and commitment, and from a scientific point of view can also result in higher quality measures.

## Discussion

In this paper we have shared some of the methodological challenges that Life Study encountered. Given its innovative and ambitious design, these challenges had been anticipated and resources allocated to tackling them. The aspects of design and analysis described in the present paper, even though ultimately they could not be followed through, are of more universal applicability and thus, we believe, potentially useful for future studies In particular, we believe that the two components, as used in Life Study, are mutually enhancing and allow causal analyses to be conducted alongside those requiring population inferences. Advances in data collection technology as well as information technology and statistical methodology now make it feasible to design and implement complex longitudinal

cohort studies that move beyond previous designs based upon national or regional populations. While in the past, the latter may have had (relative) simplicity in terms of analysis, and data collection, and been practical in terms of the technology involved, we would contend that they should no longer be viewed as the norm. In particular we consider that locally and institutionally based samples are likely to become a more important feature of large scale longitudinal studies in the future.

Of course, as we have described, embarking on a more complex design does involve a greater level of sophistication in terms of data collection and processing, and the skills and training required involved need to be taken into account. Carrying out preliminary analysis on synthetic data to ascertain how to utilise administrative data, create weights, and deal with non-response and attrition, is also important and can assist in anticipating and providing for the methodological support needed to ensure successful and full use of the resulting data.

As more administrative datasets become available from a wide variety of sources, the opportunities for complex survey designs, which allow for different data collection instruments associated with different samples whilst facilitating the combination of these samples, will increase. Although the statistical problems such sample design strategies introduce are challenging, there are sound scientific arguments to adopt such approaches.

## Acknowledgements

We are most grateful to Suzanne Thompson and John Bynner for helpful comments on an early draft. This work was partly supported by the Economic and Social Research Council (UK),

[Grant numbers ES/J007501/1, ES/L002507/, ES/L002353/1, ES/L012871/1, ES/N007549/1]. The UCL Population, Policy and Practice Programme was formed in 2014, incorporating the activities of the Centre for Paediatric Epidemiology and Biostatistics (CPEB). The CPEB was supported in part by the Medical Research Council in its capacity as the MRC Centre of Epidemiology for Child Health (award G0400546). Research at the UCL Institute of Child Health and Great Ormond Street Hospital for Children receives a proportion of the funding

from the Department of Health's National Institute for Health Research Biomedical Research Centres funding scheme.

## References


Boyd A, Golding J, Macleod J, Lawlor DA, Fraser A, Henderson J, Molloy L, Ness A, Ring S, Davey Smith G. Cohort Profile: the 'children of the 90s'--the index offspring of the Avon Longitudinal Study of Parents and Children. Int J Epidemiol. 2013 Feb;42(1):111-27. doi: 10.1093/ije/dys064.

Bynner, J. Wadsworth, M. Goldstein, H. Maughan, B. Purdon, S. Michael, R.(2007), "Scientific case for a new birth cohort study. Report to the Research Resources Board of the Economic and Social Research Council and Appendices", http://www.longviewuk.com/pages/reportsnew.shtml.

Bynner, J. Wadsworth, M. Goldstein, H. Maughan, B. Lessof, C. Michael, R.(2009), "Options for the design of the 2012 birth cohort study, Report to the Research Resources Board of the Economic and Social Research Council and Appendices", http://www.longviewuk.com/pages/reportsnew.shtml

Carpenter, J. R. and Kenward, M. G. (2013) *Multiple Imputation and Its Application*. Chichester: Wiley.

Dezateux, C; Colson, D; Brocklehurst, P; Elias, P; (2016a) 'Life after Life Study' Report of a Scientific Meeting held at The Royal College of Physicians, London, UK, 14th January 2016. (Life Study Working Papers ). Life Course Epidemiology and Biostatistics/ UCL Institute of Child Health: London, UK. http://dx.doi.org/10.14324/000.rp.1485681

Dezateux, C, Knowles, R; Brocklehurst, P, Elias, P, Burgess, S, Colson, D, et al; (2016b) Life Study Scientific Protocol. (Life Study Working Papers). Life Course Epidemiology and Biostatistics/ UCL Institute of Child Health: London, UK. http://dx.doi.org/10.14324/000.rp.1485668

Goldstein, H. (2009). "Handling attrition and non-response in longitudinal data." *International Journal of Longitudinal and Life Course Studies* **1**: 63-72.

Goldstein, H., Carpenter, J. R. and Browne, W. J. (2014), Fitting multilevel multivariate models with missing data in responses and covariates that may include interactions and non-linear terms. *Journal of the Royal Statistical Society*: Series A (Statistics in Society). 177(2), 553-564 doi: 10.1111/rssa.12022

Goldstein, H., Carpenter, J. and Kenward, M. (2016). Bayesian models for weighted data with missing values: a bootstrap approach. (submitted for publication).

Goldstein, H., Lynn, P., Muniz-Terrera, G. & Hardy, R., O'Muircheartaigh, C., Skinner, C. & Lehtonen, R. (2015). Population sampling in longitudinal surveys debate. *Longitudinal and Life Course Studies, 6, 447 – 475.* http://dx.doi.org/10.14301/llcs.v6i4.345

Kish, L. (2014). Survey Sampling. New York, Wiley.

ONS, (2013). Births statistics: Metadata. [viewed 20th January 2014]. Available from: http://www.ons.gov.uk/ons/guide-method/user-guidance/health-and-life-events/

Pearl, J. (2009). "Understanding propensity scores". Causality: Models, Reasoning, and Inference (Second ed.). New York: Cambridge University Press. *ISBN 978-0-521-89560-6*.

Wright, J., Small, N., Raynor, P., Tuffnell, D., Bhopal, R., Cameron, N., Fairley, L., Lawlor, D, A., Parslow, R., Petherick, E. S., Pickett, K. E., Waiblinger, D., & West, J. (2013). Cohort profile: The Born in Bradford multi-ethnic family cohort study. *International journal of epidemiology*, 42(4), 978-991.