This presentation was made on 26 July 2018 at the Second International Conference on the Methodology of Longitudinal Surveys, University of Essex, UK (www.mols2.org.uk)

It will form the basis of a chapter in the monograph conference book to be published by Wiley, “Advances in Longitudinal Survey Methodology”

These slides are provided for personal use only.

Please do not cite or quote without permission.

© Peter Lynn & Nicole Watson

Issues in Weighting for Longitudinal Surveys

Peter Lynn, University of Essex

Nicole Watson, University of Melbourne

An initiative by the Economic and Social Research Council, with scientific leadership by the Institute for Social and Economic Research, University of Essex, and survey delivery by NatCen Social Research and Kantar Public

Aim of the Presentation

- to outline some of the aspects of weighting that are unique to the longitudinal context

- to discuss some of the advantages and disadvantages of different possible solutions to some of the issues that are encountered

The Context

Unique features of longitudinal surveys:

1. Dynamic study population
2. Study participation dynamics (wave non-response patterns)
3. Auxiliary variables

Also:

- Multiple samples
- Longitudinal surveys as multi-purpose research resources

1. Dynamic Study Population

Population definition must include a time dimension

Many study populations (defined both by analysis aims and by data collected to date)

Population reference data (unlikely to exist)

Population entrants (representation)

Population exits (identification)

1.1 Post-Stratification

Four possible approaches:

a) Post-stratify r_1 to U_1 and thereafter rely on nonresponse adjustments;

b) Post-stratify $(r_n + i_n)$ to U_1 ;

c) Post-stratify r_n to U_n ;

d) Use admin data to construct longitudinal population targets.

where

r_n is the responding sample at wave n;

i_n is the ineligible sample at wave n (population exits);

U_n is the (cross-sectional) population at wave n;

1.1 Post-Stratification

Approach a) may be effective if no sample entrants after w1

Then, for example, analysis weight for data from w1 & w2 might be:

$$w_{i2} = dw_{i1} \times nr_{i1} \times ps_{i1} \times nr_{i2}$$

where

- dw_{i1} is design weight for element i at wave 1;
- nr_{i1} is a non-response adjustment for element i at wave 1 (based on adjusting r_1 to the characteristics of s_1);
- ps_{i1} is post-stratification adjustment for element i at wave 1 (based on adjusting the preliminarily-weighted r_1 to the characteristics of U_1 from an external source, where the preliminary weight is $dw_{i1} \times nr_{i1}$); and
- nr_{i2} is a conditional non-response adjustment for element i at wave 2 (based on adjusting r_{12} to the characteristics of (r_1 - i_2)).

1.1 Post-Stratification

Approach b) has a potential advantage over approach a) in that some post-w1 non-response errors not captured by the nonresponse adjustment may be addressed by the PS.

But rare in practice to have PS variables that could not be incorporated into the NR adjustment, so the advantage is rarely realised.

Approach b) differs from a) in that:

ps_{i1} is replaced by ps_{i2} , where ps_{i2} is PS adjustment for element i at wave n , obtained by adjusting the preliminarily-weighted (r_n + i_n) to the characteristics of U_1 , where the preliminary weight is $dw_{i1} \times nr_{i1} \times nr_{i2}$); and nr_{i2} is a non-response adjustment for element i at wave n , based on adjusting (r_{12} + i_2) to the characteristics of r_1

1.1 Post-Stratification

Approach c) post-stratifies to the most up-to-date population data available, but:

Ignores distinction between longitudinal and cross-sectional populations and therefore implicitly assumes that the sample is designed to remain cross-sectionally representative.

May be true for certain specific survey designs (but even then, only for cross-sectional analysis) but is not generally true

1.1 Post-Stratification

Approach d):

We know of no known applications (yet);

Possible if population register or other administrative data exists with complete coverage and continual updating

Population characteristics could be constructed for any longitudinal population, provided that entrance and exit dates are recorded and complete longitudinal data maintained/archived

1.2 Population Entrants

If the survey sample design includes features intended to regularly add representative samples of new population entrants (see other presentation in this session!):

a) How to derive a base weight for the sample entrants;

b) How to combine a non-response adjustment for the new entrants with one for the continuing sample.

Example scenario: household panel surveys in which all new births between waves (n-1) and n , where the mother (or possibly either parent) is a sample member, become sample members at wave n

Base Weight for New Entrants

Probability of a new birth entering the sample is product of:

a) Selection probability of the mother (parents) – which may itself be the sum of multiple probabilities;

b) Conditional probability of continuing to observe the mother (parents) until wave *n*.

Estimate of b) is typically needed to derive mother's weight, so can be used also for the new entrant. This may lead to mother's longitudinal weight (w_1-w_n) being assigned to child as base weight.

Nonresponse Adjustment for New Entrants

Will likely depend on strategy used to produce longitudinal weights for continuing sample members (see later)

If strategy involves a chain of wave-on-wave conditional NR models, then adjustment for new entrants can be made in same way, starting at wave $n+1$

If strategy involves NR models for other combinations of waves, then a dual-estimation approach of parallel models will be needed

1.3 Uncertain Eligibility

Standard non-response adjustment models assume eligibility status to be known (ineligibles are removed from the base). But…

- Identification of population exits in the sample is not always possible;
- e.g. field outcome at wave after a sample member died may be "non-contact" - particularly likely when sample member lived alone;
- Sample members may die years after last contact with survey;
- Methods needed to deal with the under-identification of mortality;
- Otherwise, estimation may be biased (respondents with similar characteristics to those who died will tend to be over-weighted)

Estimating mortality to adjust weights

Step 1:

- Estimate probability of mortality for all cases of uncertain mortality

Step 2:

- Use these estimates to adjust the weights

There are several possible approaches to each step

Step 1: Estimating mortality

(Match to death registers)

Use life tables

Explicit modelling of survival times

Implicit modelling: analogous to NR modelling, but known dead are included as “response” (rather than excluded)

(Watson, JOS, 2016)

Step 1: Estimating mortality

Use life tables:

- $\widehat{p_{ij}} = 1 - (\prod_{t=m}^{n}(1 - w_{ijt}d_{jt}))$ - individual method
- $\widehat{p_{ij}} = \frac{n_j(1-\prod_{t=1}^{n}(1-w_{jt}d_{jt}))-o_{jn}}{n_j-c_{jn}-o_{jn}}$ - group method

where:

Element *i* in subgroup *j* (e.g age x sex) was last observed (alive) in year *m*;

d_{jt} is mortality rate for individuals in subgroup *j* in year *t*;

w_{ijt} are weights reflecting the relevant proportion of a calendar year

n_j is initial sample size in group *j*, of which c_{jn} known alive in year *n*;

o_{jn} is observed number of deaths in group *j* by year *n* (year of current wave)

Step 2: Adjusting weights

Impute dichotomous alive/dead indicator,

e.g. $a_{ij} = 1 \; if \; r_{ij} > \widehat{p_{ij}}; \quad r_{ij} \sim U(0,1)$ - then usual NR model

Usual NR model and then scale weights for each group *j by* k_j:

$$k_{jn} = \frac{c_{jn} + (1 - \widehat{p_{ij}})u_{jn}}{c_{jn} + u_{jn}}$$

Usual NR model, but with cases weighted by $\left(1 - \widehat{p_{ij}}\right)$

Case Study: BHPS

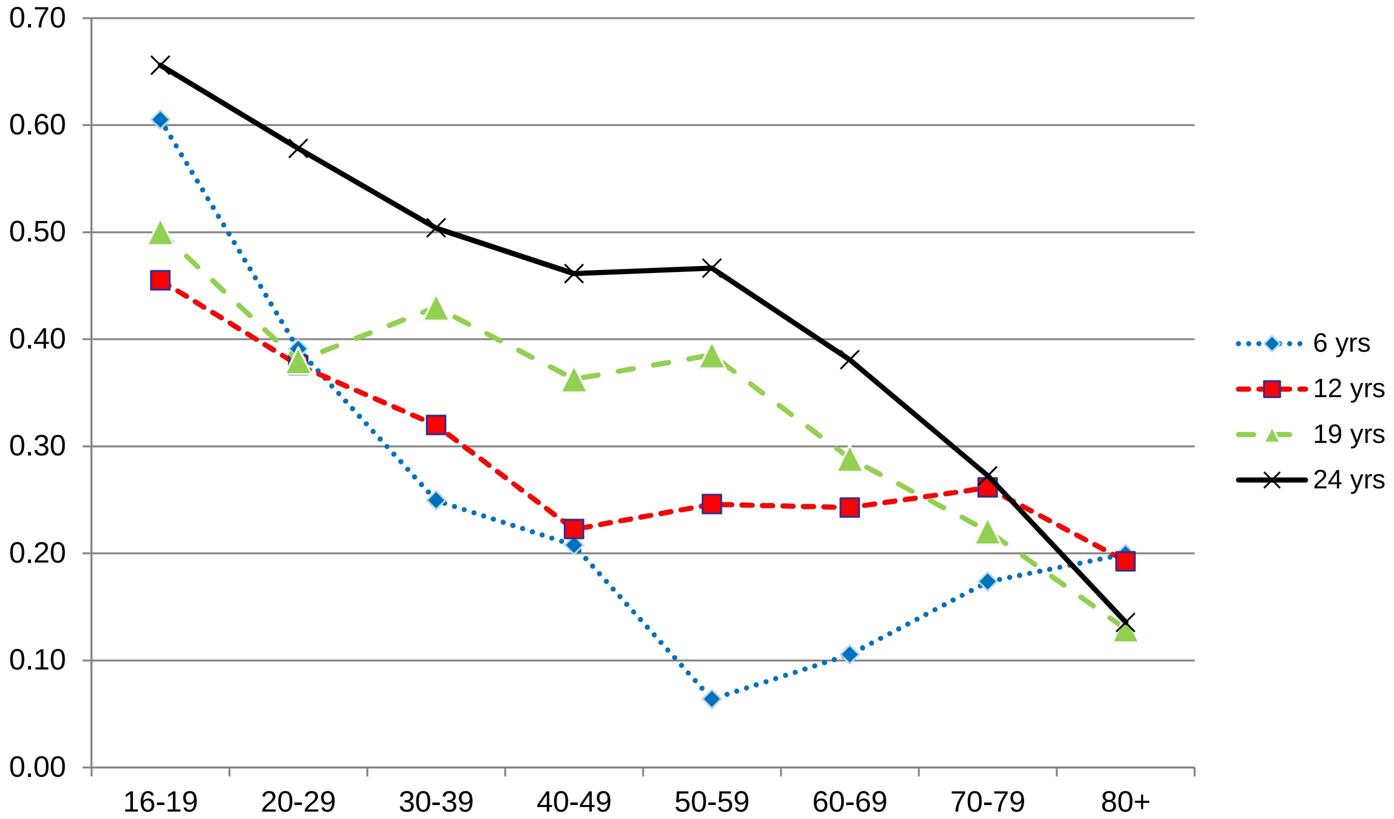
For each age x gender group, for each survey wave:

o_{jn} - observed number of deaths in group *j* by year *n*

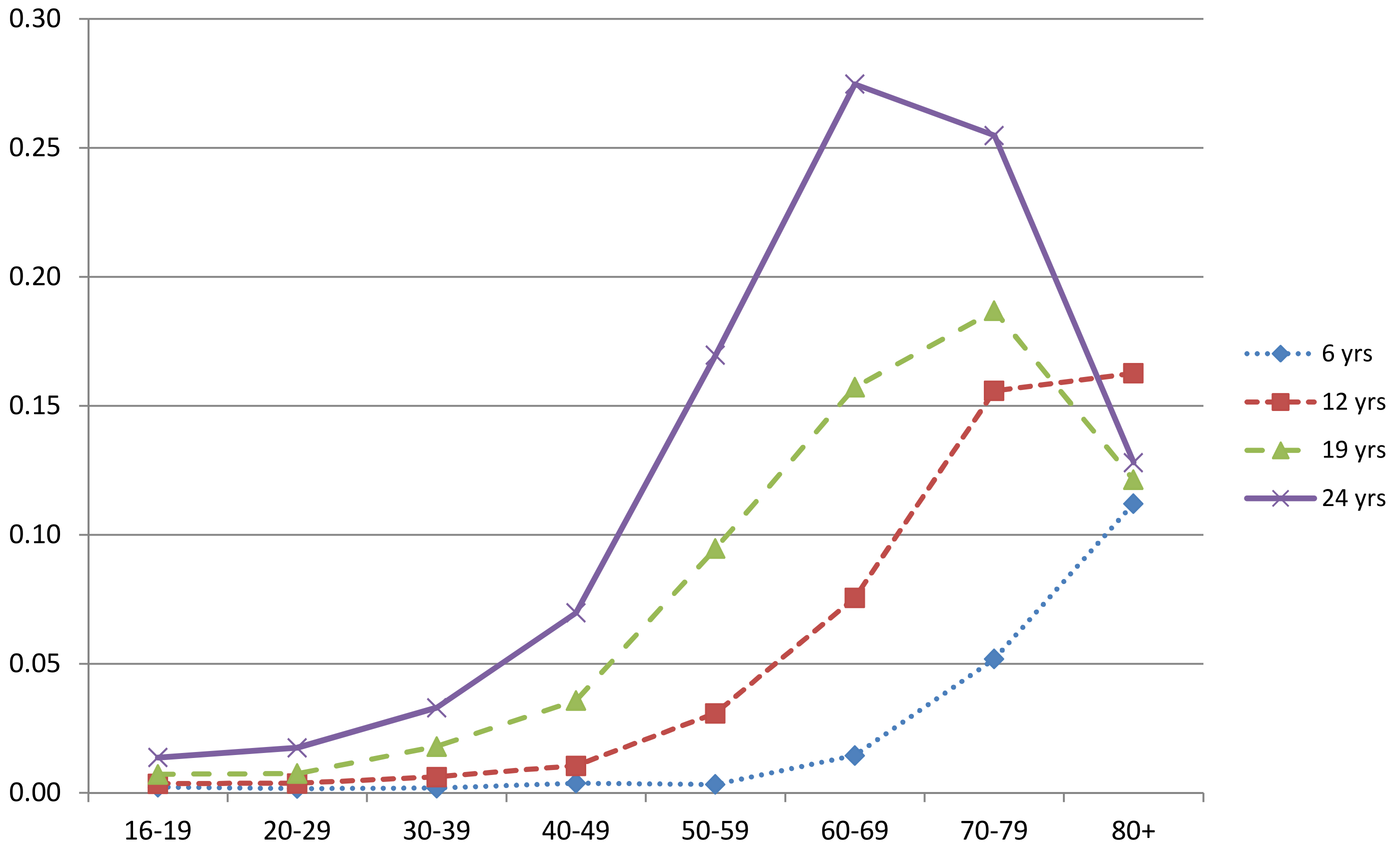
$e_{jn} = n_j \left(1 - \prod_{t=1}^{n}\left(1 - w_{jt} d_{jt}\right)\right)$ - expected number

Proportion of deaths not identified: $1 - \frac{o_{jn}}{e_{jn}}$

Proportion of deaths not identified, by age at wave 1 and elapsed years, BHPS



Deaths not identified as a proportion of sample, BHPS



2. Sample Participation Dynamics

Sample available for longitudinal analysis depends on wave non-response patterns

Sample can differ markedly between different combinations of waves, both numerically and in terms of composition

After n waves, there are 2^n-1 possible wave combinations

If i instruments per wave, 2^{in}-1 possible combinations

e.g. n=10, i=2: $2^{in}-1 > 1$ million

Impractical to produce weights for each combination

Possible approaches

- Provide weights for a (necessarily small) subset of the possible combinations of instruments/waves
- Provide weights for each pair of instruments/waves
- Provide means for users to produce analysis-specific weights

2.1 Weights for subsets of combinations

Which combinations? Possible criteria (Lynn & Kaminska, 2010):

- Analytic use
- Levels of non-response
- Correlates of non-response
- Impact on estimates

In reality, we tend to use heuristics based largely on the first two criteria

2.1 Weights for subsets of combinations

What to do when desired analysis base is not one of those for which weights have been produced:

a) Use the weight from the hierarchically-superior combination that is closest in sample size to the analysis base

b) Adjust the weight from a hierarchically-superior combination

c) Use the weight from a hierarchically-inferior combination that is close in sample size to the analysis base.

2.1 Weights for subsets of combinations

a) Use the weight from the hierarchically-superior combination that is closest in sample size to the analysis base.

e.g. if analysis base is combination (D, E, F, G), use weight for the combination of three of those four that has smallest sample size.

In practice, a weight may have been produced for only one of the four possible 3-instrument combinations, say (D, E, F), in which case that weight should be used.

All sample units in the analysis sample will have a non-zero weight.

The weight will be sub-optimal in so far as relevant correlates of response to $G|(D, E, F)$ differ from relevant correlates of response to (D, E, F).

2.1 Weights for subsets of combinations

b) Adjust the weight from a hierarchically-superior instrument combination

user fits their own model of response to *G* conditional on response to (D, E, F)

predicted values from this model form adjustments to the weight for (D, E, F)

users should be advised on how to select covariates for the model and how to save the predicted values and use them as weight adjustments

2.1 Weights for subsets of combinations

c) Use the weight from a hierarchically-inferior instrument combination that is close in sample size to the analysis base

If a weight is available for the combination (D, E, F, G, H) and there are very few sample members who responded to (D, E, F, G) but not to H, then it may be appropriate to use the weight produced for (D, E, F, G, H).

This will result in a small number of potential members of the analysis base being excluded (as they will have a weight of zero), but the nonresponse correction will be appropriate as the *de facto* analysis base is (D, E, F, G, H)

a) *vs.* c) involves a bias-variance trade-off

2.2 Weights for each pair of instruments

Provide weights for baseline participating sample

And for every subsequent combination of two instruments (not only ‘consecutive’ pairs)

If *N* instruments, total number of sets of weights to produce will be $1+\sum_{n=1}^{N}(n-1)$, which is very considerably fewer than the 2^N-1 possible combinations.

e.g. if $N=10$, then $1+\sum_{n=1}^{N}(n-1)=46$, whereas $2^N-1=1{,}023$.

2.2 Weights for each pair of instruments

For each pair, weight should correspond to the inverse of the probability of responding to one of the instruments conditional upon having responded to the other

e.g. $w_{i(D|A)} = 1/\hat{P}(I_{iD} = 1|I_{iA} = 1)$, where I_{in} is a 0/1 indicator of whether sample member i responded to instrument n.

Logical direction of conditionality will often correspond to the temporal ordering of the instrument administration – but this is not necessary

2.2 Weights for each pair of instruments

Then, users can derive a weight for any combination by taking the product of component weights.

e.g. weight for analysing (A, B, F, M) could be computed as $w_{i(ABFM)} = w_{i(A)} \times w_{i(B|A)} \times w_{i(F|B)} \times w_{i(M|F)}$.

If data structure strictly hierarchical (e.g. data collection protocol is to administer instrument B only if A is completed, etc) then this is equivalent to explicitly producing weights for each possible combination of instruments using chained conditional NR models.

But if structure is non-hierarchical, weights will be suboptimal to the extent that some of the conditional nonresponse models differ from the optimal ones. *cf.* $w_{i(F|B)}$ *vs.* $w_{i(F|AB)}$

2.3 Analysis-specific weights

Systems to produce weights on demand for any combination of instruments (or possibly any sub-sample)

Example: the National Longitudinal Surveys of Youth (NLSY) in the USA has a custom weighting program (www.nlsinfo.org/weights)

Program applies a simple algorithm to calculate a post-stratification adjustment to the wave 1 weight, based on population information from the time the sample was selected

Ignores: mortality, population dynamics, survey data as auxiliary vbls

Would be challenging to produce a system sufficiently sophisticated to deal with all, or even most, of the issues relevant to weighting for longitudinal surveys

3. Auxiliary Variables

A big strength of longitudinal surveys: powerful auxiliary data for all forms of missing data post-wave 1.

However, there is a balance to be struck between:

- Harnessing the power of these auxiliary data
- Keeping the weighting process manageable and parsimonious

Approaches

a) Explicitly model NR for each combination in a single step and apply a single adjustment

b) Model each step in the attrition process sequentially and apply a series of multiplicative adjustments

c) Adopt a hybrid approach, in which some steps in the attrition process may be modelled separately, while others may be combined into a single model

Approaches

Same set of models for all member of the analysis sample:

a) Explicitly model NR for each combination in a single step and apply a single adjustment

b) Model each step in the attrition process sequentially and apply a series of multiplicative adjustments

c) Adopt a hybrid approach, in which some steps in the attrition process may be modelled separately, while others may be combined into a single model

Varying models:

d) Chain of models can differ between sample subsets, to make fuller use of available covariates, provided each chain is comprehensive and hierarchical

3.1 Single-step model for each combination

Requires auxiliary variables available for all sample members

Auxiliary variables therefore restricted to sampling frame, observation, and externally-linked variables

Each time weights are needed for a new combination (for example, because a new wave has been carried out), all previously-produced weights are ignored and development of a weighting model begins again from scratch

3.2 Model each step in the process

Maximises the extent and relevance of auxiliary variables that can inform the weights, as these can include all survey measures from all previous steps in the process

Each time weights are needed for a new combination, the starting point is to use the previously-produced weight as a ‘base weight’ and to develop a new adjustment to be made to that weight, based on a model of conditional response to the new wave/ instrument

3.3 Hybrid approach

Some steps in the attrition process may be modelled separately, while others may be combined into a single model

e.g. model wave 1 nonresponse using frame data as auxiliary variables, but then model the combination of wave 2, 3 and 4 nonresponse in a single step, using wave 1 survey data as auxiliary variables

3.4 Varying models

e.g. weights for (1,4) from chain of models:

(1), (4|1) - for w2 nonrespondents

(1), (2|1), (4|1,2) - for w2 respondents

Allows w2 data to be used as auxiliary data for w2 respondents

But note the requirement for hierarchical models, so <u>all</u> wave 1 respondents are included in the models (4|1) and (2|1) above, not only those who will receive their weight via that chain

Conclusions

Overall weighting strategy is a combination of many decisions regarding how to deal with each of the issues outlined

These decisions are often based on heuristics / intuition. There is little or no research into the properties of alternative approaches and little or no development of evaluation methods

Future Agenda

Exploiting administrative data to construct longitudinal population targets

Evaluation of alternative methods for handling uncertain mortality, especially the second step (weight adjustment). & extension to emigration or other categories of ineligibility

Research into factors that determine the extent and nature of suboptimality in using weights for hierarchically-superior or hierarchically-inferior combinations

Development of the instrument-pair method

Evaluation in a variety of context (patterns/levels of nonresponse, mortality, etc)

Issues in Weighting for Longitudinal Surveys

Peter Lynn, University of Essex

Nicole Watson, University of Melbourne

An initiative by the Economic and Social Research Council, with scientific leadership by the Institute for Social and Economic Research, University of Essex, and survey delivery by NatCen Social Research and Kantar Public