Appendix B. Sampling errors

B1 Sources of error in surveys

As with any survey, estimates resulting from the Infant Feeding Survey are subject to different sources of error. Sampling theory describes two sources of error: *systematic error* (or bias) and *random error.*

Systematic error or bias arises when respondents to the survey are not representative of the population of interest. This can arise if either the original selected sample was unrepresentative or if the response rate to the survey is low and varies significantly across different groups of participants. A rigorous sample design with a sample selected from an exhaustive sampling frame will eradicate the former, while high response rates across all respondent groups will resolve difficulties of the latter type.

The sample for the 2010 Infant Feeding Survey was drawn from all live births registered during a defined sampling period in August and October 2010. In Wales and Northern Ireland all mothers giving birth in this period were selected to be part of the study, while in England and Scotland a random sample of mothers were selected.

Overall, the main source of error is sampling error. The extent of this depends on the natural variation in any measure that is collected and the sample size achieved. The following formulae and discussion explain how to estimate the sampling error around the survey estimates.

B2 Standard errors and confidence limits

For any percentage estimate p, the 95% confidence interval for the population statistic P is:

$$p \pm 1.96 * se(p) \qquad \text{(A)}$$

where se(p) is the standard error of the estimate

The standard error of p is unknown but for a Simple Random Sample (SRS) survey estimate it is estimated by the quantity:

$$\sqrt{p(1-p)/n} \qquad \text{(B)}$$

where *n* is the sample size for a particular group

This SRS formula applies to a sample and sub-groups when the design is a random selection, the sample size is a small proportion of the population, and respondents have equal selection probabilities. However, the 2010 Infant Feeding Survey sample design differed from a SRS in two ways.

- In all four countries, a high proportion of the population of mothers giving birth in August-October 2010 was selected and took part in the survey; this by itself would mean that the survey estimates have smaller sampling errors than those suggested by the formula.

- There were unequal selection probabilities in England, Scotland and also overall, for UK estimates. Within England and Scotland, where only a sample of mothers was selected, mothers living in the most deprived quintile of each country's Index of Multiple Deprivation were sampled at a higher rate. Overall the whole sample was skewed towards Wales, Scotland and Northern Ireland, meaning that UK estimates will also be affected by unequal selection probabilities. Also within all countries, although non-response was relatively low, it was higher amongst certain groups so weighting was applied to try to reduce any bias arising from this. This factor by itself would mean that the survey estimates have larger sampling errors than those suggested by the formula.

The effect of these two differences means that the formula for the standard error is not correct and an adjustment factor, called a design factor, needs to be calculated to account for these differences. The design factor is the ratio of the standard error of the complex design to that of the standard error of a simple random sample of the same size.

Put another way, the design factor (deft) is the factor by which the standard error of an estimate from a simple random sample has to be multiplied to give the true standard error of the complex design (this is referred to as the complex standard error or the true standard error).

A design factor of less than one arises when a sample is more precise and has smaller standard error than a SRS, while a design factor of greater than one arises when a sample has standard errors larger than those that would be obtained from a SRS of the same size.

The complex standard errors and design factors for selected Infant Feeding Survey measures are given in the following tables B.1-B.5 together with unweighted and weighted bases for selected measures in each country and for the United Kingdom. The weighted bases are given simply as an indication of the sample size and to be consistent with the conventions used throughout the rest of the report. However, the calculation of design standard error and design factor are always based on the unweighted sample size for any particular sub-group.

Design factors in Wales, Scotland and Northern Ireland are smaller than for a SRS since a high proportion of mothers were surveyed, while design factors for England are around the same as for a SRS. Design factors for UK estimates are higher than those in the individual countries due to the skew in the sample towards the smaller countries.

As an illustration the following explains how to calculate a confidence interval for the incidence of breastfeeding figure for the UK quoted in Table B.1.

Characteristic	Sample sub-group	Percentage	Complex Standard Error	Design factor	Weighted base
Incidence of breastfeeding	UK	81%	0.376	1.202	15722

Using the above figures, the 95% Confidence Interval is then calculated as

81% $\pm$ 1.96*0.376 = 80.3% to 81.7%

It is important that complex standard errors or design factors are used for confidence intervals and significance tests for the UK since the complex standard errors are generally greater than those produced by the SRS formula. If the SRS formula is used, significant differences may be found erroneously.

For survey estimates that are not included in tables B.1-B.5, the 95% confidence intervals for a percentage p can be calculated using the formula:

$$p \pm 1.96*deft*se\,(p) \qquad (C)$$

Where se(p) is the standard error from equation (B) assuming a simple random sample, and the 'deft' is the design factor. For measures not shown in tables B.1 – B.5, a good estimate of the standard error can be obtained by using the largest design factor shown for the relevant country from Tables B.1 – B.5.