

FAB data analysis plan (objective 2)

Agreed by co-investigators June 2012

Note added 06.02.13

This data analysis plan is attached as a supplementary file for the paper submitted to the BMJ “Factors associated with breastfeeding in England: an analysis by primary care trust”. The submitted paper covers the analysis of objective 2a as described in this analysis plan. Anything in this analysis plan which specifically relates to objective 2b should be ignored.

1. Aims and objectives

The overall objectives of the FAB project as described in the project protocol are to:

1. Collate area-based data on breastfeeding prevalence at 6-8 weeks, socio-demographic factors and breastfeeding interventions
2. Use these data to identify predictors of variation between areas in breastfeeding prevalence at 6-8 weeks
3. Use individual level data to measure the demand for breastfeeding services
4. Monitor changes over time in breastfeeding prevalence and interventions, and evaluate the implementation of any subsequent changes in service.

This data plan covers the analyses planned to address objective 2 (to use data to identify predictors of variation between areas in breastfeeding prevalence). We will address the following specific objectives as part of objective 2:

- a. To identify socio-demographic predictors of variation between areas in breastfeeding rates
- b. To measure the effect of a specific breastfeeding intervention (the Baby Friendly Initiative) on area-based breastfeeding rates
- c. To measure the effect of other indicators of breastfeeding support on area-based breastfeeding rates

The first two objectives (objectives 2a and 2b) are covered in this analysis plan. The inclusion of other indicators of breastfeeding support (objective 2c) will be agreed after primary data collection has been planned.

2. Design

This is an ecological (area-based study) study making use of routine aggregate data. A subsequent phase may involve primary data collection and will address the third objective (objective 2c) of this study.

3. Variables

3.2 Definition of outcomes

The primary outcome is breastfeeding prevalence (‘any breastfeeding’) at 6-8 weeks. Secondary outcomes are the prevalence of exclusive (‘total’) breastfeeding at 6-8 weeks and initiation of breastfeeding

Information on breastfeeding status at 6-8 weeks (breastfeeding prevalence) is collected by the GP or HV conducting the routine 6-8 week infant check and reported by PCTs to the Department of Health (DH) as part of the of the Vital Signs Monitoring Return programme (VSMR). Infants exclusively breastfed are those who are receiving breast milk and “NOT receiving formula milk, any other liquids or food”. The proportion ‘any breastfeeding’ comprises those infants

who are totally breastfed or who are receiving some breast milk in addition to other milk, liquids or food. The denominator for this outcome is the number of infants due a 6-8 week check.

Data on breastfeeding initiation is typically collected by midwives in acute trusts and again forms part of the VSMR. In this case, breastfeeding initiation is defined as the “mother...having initiated breastfeeding if, within the first 48 hours of birth, either she puts the baby to the breast or the baby is given any of the mothers breast milk”. The denominator for this outcome is the number of maternities.

A third secondary outcome – drop off in breastfeeding between initiation and 6-8 weeks – may also be included.

Data on breastfeeding prevalence at 6-8 weeks and breastfeeding initiation are reported quarterly by PCTs. All outcome data used in this analysis will relate to the time period April 2010-March 2011 (2010-11 quarters 1-4).

The DH releases PCT level figures where the data pass validation checks and meets a minimum level of data coverage. Coverage is defined as the percentage of infants due a 6-8 week check for whom a breastfeeding status was recorded (breastfeeding prevalence), or as the percentage of maternities for which an initiation status was recorded (breastfeeding initiation). DH requires coverage to be a minimum of $\geq 90\%$ (breastfeeding prevalence quarters 1-3) or $\geq 95\%$ coverage (breastfeeding initiation, breastfeeding prevalence quarter 4 only). PCTs will therefore only be included in the analysis if they meet these criteria.

3.3 Definition of socio-demographic factors (objective 2a)

The following potential socio-demographic factors have been identified and will be included in the analysis as appropriate: area-based deprivation, ethnicity, maternal age and maternal smoking. All are available at the PCT level.

Deprivation

The area-based deprivation indicator to be used is the Index of Multiple Deprivation (IMD) 2010. This index measures a broad concept of deprivation and is derived from census variables and other more recent data sources. A total of 38 different indicators are aggregated into seven domains: income; employment; health and disability; education, skills and training; barriers to housing and services; living environment; and crime. These indicators are weighted and combined to calculate a final IMD ‘raw’ score. A high score indicates greater deprivation. The IMD is calculated at Lower Super Output Area (LSOA), of which there are 32,482 in England. This analysis makes use of a dataset which reports IMD 2010 score at the PCT level.

Ethnicity

Two different ethnicity variables have been identified for use, both reflecting the proportion from a Black and Minority ethnic (BME) background, defined in this case as non-White British. The first measure estimates the percentage of the *overall* PCT population from BME backgrounds (PEEG - Population Estimates by Ethnic Group). The estimate is derived from the 2001 Census and is calculated by the Office for National Statistics (ONS) using a cohort component methodology taking into account births, deaths, and migration to and from the area. Estimates for 2009 are the most recently available figures and are used in this analysis. The second measure summarises the proportion of women from a BME background who delivered in the given time period (2010-11). These figures are taken from HES data. Although the latter variable is most pertinent as it relates to the maternity population, there is some concern about the level of missing data. For this reason we will include both variables in descriptive analysis, and will adjust for the one which changes the effect measures the most, provided the level of missing data or accuracy of data is not an issue.

Maternal age

We will include two indicators of births by maternal age as covariables: the percentage of mothers aged <20 and the percentage of mothers aged ≥ 35 . These data are drawn from HES delivery episode data and are available from ChiMat at PCT level for the year 2009-10.

Smoking

Smoking status is collected at the time of delivery and is reported as the percentage of women giving birth who are current smokers at the time of delivery. This is another data item included in the VSMR and the data for 2010-11 are used in this analysis.

3.3 Definition of the Baby Friendly Initiative (objective 2b)

The explanatory factors for objective 2b are Baby Friendly Initiative (BFI) status in the hospital (acute trust) and BFI status in the community.

BFI status comprises of multiple categories. As hospitals or community organisations move through the pathway to full accreditation they pass through the following milestones and awards: register of intent, certificate of commitment, stage 1, stage 2, before finally achieving full accreditation (stage 3).

BFI status will be measured at April 2010 to reflect status at the beginning of the period of outcome measurement. In the vast majority of cases, community BFI status relates directly to the same geographical area (PCT) used in the collection of breastfeeding data, reflecting the same unit of analysis. Occasionally, BFI accreditation relates to a specific provider arm rather than general services.

Hospital BFI status will need to be mapped to PCT level outcome data to enable us to measure the effect of hospital accreditation on breastfeeding rates. To facilitate this, data on the provider of maternity care by PCT of responsibility has been sought from HES. Where multiple acute trusts deliver maternity care to a single PCT population, an algorithm has been developed to take into account the proportion of deliveries attributable to each provider within a PCT. Using this it is possible to estimate the number of deliveries in each PCT taking place in a unit with each level of BFI award.

4. Data management

4.1 Breastfeeding outcomes

Data on breastfeeding outcomes are reported by quarter. Annual figures for 2010-11 will be calculated by summing the relevant quarterly figures and calculating the mean across the period. These figures will only be calculated for PCTs contributing data of an acceptable quality (i.e. meeting DH validation checks) for at least two of the four quarters in 2010-11.

3.2 Socio-demographic variables

All of the socio-demographic variables in this analysis are continuous variables. In addition to presenting summary data (means, ranges etc.), data may be grouped for the purpose of analysis. Continuous variables will be transformed into ordered categorical variables using quintiles or quartiles, or well-defined cut-offs if their association with the outcome of interest is not linear.

3.3 BFI status

BFI will be included in descriptive analysis as an ordered categorical variable with six groups. For regression analysis, BFI status will be collapsed into two or three groups as detailed in Tables 2 and 3. These groupings were agreed following advice from BFI staff and Advisory Group members. Time since award will be considered for the longer established hospital award but not for community BFI status as this is a more recent award.

5. Analysis plan

5.1 Descriptive analysis (objectives 2a and 2b)

Descriptive analysis will involve an examination of data quality and completeness, Crude breastfeeding rates will be reported and summarised. PCTS will be described with respect to each of the variables included in the analysis. This will involve the presentation of summary tables, scattergrams and other visual displays. Prevalence estimates for the primary and secondary outcomes will be presented alongside confidence intervals. All the potential socio-demographic indicators being considered are continuous variables, and as such, means and standard deviations will be presented where distributions are approximately normal. For variables with a non-normal distribution or those with extreme values, the median and interquartile range will be presented. Frequencies and percentages will be reported for the explanatory variables and for grouped continuous variables.

5.2 Multivariable analysis (objectives 2a and 2b)

All statistical analyses will be conducted using Stata version 11. All tests will be two tailed and a 5% significance level will be used unless specified otherwise.

Statistical methods

Logistic regression will be used to estimate the effect of socio-demographic variables and breastfeeding support on breastfeeding outcomes. Aggregated data will be modelled as individual data and random effects models will be used to take into account the clustered hierarchical nature of the data.

Variables will only be retained in models where there is evidence of an independent association. This will be assessed by entering all potential explanatory variables in a regression model, dropping the least significant variable one by one, and examining the model as each variable is dropped until all variables remaining in the model are associated ($p < 0.05$) with the outcome. This strategy will be repeated for each relationship under study.

For objective 2a, an adjusted odds ratio (OR) will only be presented where the socio-demographic variable is associated with the outcome in crude analysis at $p < 0.10$ (i.e. looks to be a 'predictor' of breastfeeding outcomes).

Where adjusted odds ratios are reported, a minimum of three sets of odds ratios (OR) will be presented for each specific analysis: i) an unadjusted OR for all PCTs with valid outcome data, ii) an unadjusted OR for all PCTs with valid outcome data and no missing data for any explanatory factor, iii) an adjusted OR for all PCTs with valid outcome data and no missing data for any explanatory factor. For analyses undertaken for objective 2b, a fourth OR will be presented. This will be adjusted for socio-demographic variables, BFI status of hospital/community (whichever is not the main exposure), and (only where the outcome is breastfeeding at 6-8 weeks) breastfeeding initiation.

For objective 2a (identifying socio-demographic predictors), collinearity will be checked using summary tables showing the association between pairs of variables and by looking at the stability of coefficients and standard errors in models which include 'correlated' variables. Where extreme collinearity is present, only the strongest variable (as assessed using p values) will remain in the model.

For objective 2a, area-based deprivation and ethnicity will be considered as potential effect modifiers, and their role will be examined using Forest plots and tests for heterogeneity.

Planned analyses

Objective 2a

For this objective, we will examine the relationships detailed in Table 1.

Table 1. Analyses planned for objective 2a (socio-demographic predictors of breastfeeding rates)

Exposure	Outcome
Deprivation	Breastfeeding initiation
	Any breastfeeding at 6-8 weeks
	Exclusive breastfeeding at 6-8 weeks
Ethnicity	Breastfeeding initiation
	Any breastfeeding at 6-8 weeks
	Exclusive breastfeeding at 6-8 weeks
Young maternal age	Breastfeeding initiation
	Any breastfeeding at 6-8 weeks
	Exclusive breastfeeding at 6-8 weeks
Older maternal age	Breastfeeding initiation
	Any breastfeeding at 6-8 weeks
	Exclusive breastfeeding at 6-8 weeks
Maternal smoking	Breastfeeding initiation
	Any breastfeeding at 6-8 weeks
	Exclusive breastfeeding at 6-8 weeks

Objective 2b

Table 2 lists the planned analyses for the investigation of hospital BFI status. For the analysis looking at the effect of hospital BFI status on breastfeeding, the analysis is complicated by the fact that BFI status is not a simple categorical variable. There may be more than one provider of maternity services for each PCT, so instead of having single hospital BFI status for each PCT, we will instead model the percentage of births at a facility with each level of BFI award. BFI status is represented as 6 non-independent values where the 6th value is determined by the other 5 (since the sum of all values = 100). For example, assume a record for a single PCT ("PCT 1") is as follows:

	% of births at					
	No info	Register of intent	Certificate of commitment	Stage 1	Stage 2	Stage 3 (full)
PCT 1	25	5	10	35	15	10

For analysis A, we will include in the model a variable indicating the % of births in a stage 2/3 hospital, and a variable indicating the % of births in a hospital with no information/intent. Using the example above, the figures for this PCT would be 25% (15+10) and 30% (25+5) respectively. This is similar to our approach looking at maternal age, where we also plan to include in the model only the % of births in the 1st and 3rd age groups (three age groups in total). For analysis E, we would include only one variable, indicating the % of births at a facility with a status other than full accreditation. Using the example above, this value would be 90% (25+5+10+35+15).

Table 2. Analyses planned for the effect of hospital baby friendly status on breastfeeding rates (objective 2b)

Exposure		Categorisation	Outcome
Hospital BFI status	A	1. No info/register of intent 2. Certificate of commitment/stage 1 3. Stage 2/full	Breastfeeding initiation
			Any breastfeeding at 6-8 weeks
			Exclusive breastfeeding at 6-8 weeks
	B	1. No info/register of intent 2. Certificate of commitment/stage 1/stage 2 3. Full	Breastfeeding initiation
			Any breastfeeding at 6-8 weeks
			Exclusive breastfeeding at 6-8 weeks
	C	1. No info/register of intent 2. Certificate of commitment/stage 1/stage 2/full	Breastfeeding initiation
			Any breastfeeding at 6-8 weeks
			Exclusive breastfeeding at 6-8 weeks
	D	1. No info/register of intent/certificate of commitment/stage 1 2. Stage 2/full	Breastfeeding initiation
			Any breastfeeding at 6-8 weeks
			Exclusive breastfeeding at 6-8 weeks
	E	1. No info/register of intent/certificate of commitment/stage 1/stage 2 2. Full	Breastfeeding initiation
			Any breastfeeding at 6-8 weeks
			Exclusive breastfeeding at 6-8 weeks

Table 3 lists the planned analyses looking at the effect of community baby friendly status. This analysis is straightforward as there is a single community BFI status for each PCT.

Table 3. Analyses planned for the effect of community baby friendly status on breastfeeding rates (objective 2b)

Exposure		Categorisation	Outcome
Community BFI status	A	1. No info/register of intent 2. Certificate of commitment/stage 1 3. Stage 2/full	Breastfeeding initiation
			Any breastfeeding at 6-8 weeks
			Exclusive breastfeeding at 6-8 weeks
	B	1. No info/register of intent 2. Certificate of commitment/stage 1/stage 2 3. Full	Breastfeeding initiation
			Any breastfeeding at 6-8 weeks
			Exclusive breastfeeding at 6-8 weeks
	C	1. No info/register of intent 2. Certificate of commitment/stage 1/stage 2/full	Breastfeeding initiation
			Any breastfeeding at 6-8 weeks
			Exclusive breastfeeding at 6-8 weeks
	D	1. No info/register of intent/certificate of commitment/stage 1 2. Stage 2/full	Breastfeeding initiation
			Any breastfeeding at 6-8 weeks
			Exclusive breastfeeding at 6-8 weeks
	E	1. No info/register of intent/certificate of commitment/stage 1/stage 2 2. Full	Breastfeeding initiation
			Any breastfeeding at 6-8 weeks
			Exclusive breastfeeding at 6-8 weeks

Treatment of missing data

The percentage of missing data for the outcome variables will not exceed 10%. There should be minimal, if any, missing data for BFI status. If the level of missing data for covariates exceeds $\geq 10\%$ we will explore strategies to address missing data e.g. multiple imputation.