Estimating fertility by ethnic group (Paul Norman)

## The data shopping list …

National level
- Births by age of mother, detail by 5 year age-group or 'better'
- Populations at risk by matching age detail
- Large scale survey microdata with age, sex, 'subgroup' (ethnicity, religion, race, country of birth) & indicators of presence of dependant child (aged < 5) & 'generation' or 'relationship'

Subnational level
- Births by age of mother, detail by 5 year age-group or 'better'
- Populations at risk by matching age detail
- Populations by ethnic group by 5 year age-group

Vital Statistics
- Geography: national to local
- Age: five year detail
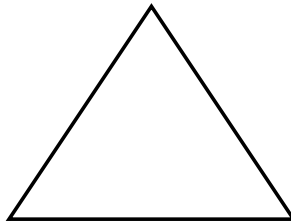
Populations
- Geography: national to local
- Age: five year detail
- Ethnicity: up to 16 groups (2001)

Survey
- Geography: national (& region)
- Age: (SYA &) five year detail
- Ethnicity: (15,) 8 & 4 groups (2001)

Remember that ethnic groups may not match between sources in definition and in the feasibility through reliability and small number problems.

Survey	White	Black		South Asian			Chinese & Others	
Intermediate	White	Black-Caribbean	Black-African	Indian	Pakistani	Bangladeshi	Chinese	Others
Census								

The recipe …

a.) National and local indicators: Vital Statistics and population-based fertility rates

      Age-Specific Fertility Rates and Total Fertility Rates

      CWR-derived TFRs by ethnic group

b.) National level indicators: Using a large scale survey to estimate fertility by ethnic group

Developing the dataset

      Defining and selecting a study population

      Variable construction

Some basic analyses to indicate fertility levels by ethnic group

      CWR-derived TFRs by ethnic group

Some more advanced techniques to indicate fertility levels by ethnic group

      1.) Statistical modelling: binary logistic regression

      2.) Mathematical function: Hadwiger curve

c.) Triangulation

      Combining the evidence

a.) National and local indicators: Vital Statistics and population-based fertility rates

The file 2001-births.xls has the birth counts by age of mother (five year groups) for each local authority in England & Wales and the national level counts.

The file 2001-population.xls has counts of women of childbearing age by age-group (by all women and by eight ethnic groups.

Calculate Age-Specific Fertility Rates and Total Fertility Rates for all women at national and local authority levels.

      Age-Specific Fertility Rates (ASFRs) = births / women * 1,000 for each age-group

      Total Fertility Rates (TFRs) = sum of ASFRs / 1,000 * age interval

Calculate Child : Woman Ratios (CWRs) by ethnic group at national and local authority levels and use these to estimate TFRs by ethnic group

      $TFR(e) = TFR(aw) * (CWR(e) / CWR(aw))$

Consider a population threshold so that where there is a 'small' population by ethnic group, the CWR is replaced by the national CWR for an ethnic group or an average of the national ethnic specific CWR and the local CWR (thereby capturing ethnic and local dimensions of fertility). (Later you might use the survey data to adjust the CWRs to better link mothers and children.)

b.) National level indicators: Using a large scale survey to estimate fertility by ethnic group

The Labour Force Survey (LFS) is a large scale government survey of persons in households living at private addresses in Great Britain. Its purpose is to provide information on the labour market which can then be used to develop, manage, evaluate and report on labour market policies. It is conducted by the Office for National Statistics.

Whilst the LFS has 'labour' as its main focus, the richness of variable information included means that the source can be used to research other aspects. Since information is in the LFS on age, sex and ethnic group along with indicators of household structure including the presence of dependant children.

The Labour Force Survey for 2001 includes data (over 600 variables) on nearly 140,000 respondents of whom around 25,000 are females of childbearing age.

The aim here is to calculate indicators of fertility by ethnic group. These could be:
• Child : Woman Ratios used in conjunction with Vital Statistics-derived Total Fertility Rates
and / or
• Age-Specific Fertility indicators from which to derive Age-Specific Fertility Rates

The data preparation stage involves:
• Selecting a study population
• Determining whether a child is present
• Aggregating ethnic groups where numbers are small

Defining and selecting a study population

Women aged 15 to 44 by ethnic group (using five year age-groups)

• Who are not (as far as it is possible to tell) the sister, grandmother; or

• Otherwise not the mother of a child in the family or household.

*SPSS syntax to extract a study population.

*Yes, this could be achieved in one step! .

*Step 1 .

*Select women of childbearing years.

```
USE ALL .
COMPUTE filter_$=(sex = 2 & age > 14 & age < 45) .
VARIABLE LABEL filter_$ 'Women of childbearing years' .
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected' .
FORMAT filter_$ (f1.0) .
FILTER BY filter_$ .
EXECUTE .
```

*When happy with the subpopulation, delete the unwanted cases .

*Syntax which deletes unwanted cases .

```
USE ALL.
SELECT IF (sex = 2 & age > 14 & age < 45) .
EXECUTE .
```


*Step 2 .

*Allow for women unlikely to be the mother of a child who may be present .

* Use the variable **relhrp** "Relationship to household reference person" .

*Syntax which filters.

```
USE ALL .
COMPUTE filter_$=( (relhrp > -1 & relhrp < 4 ) or (relhrp = 20) ) .
VARIABLE LABEL filter_$ 'Women likely to be the mother if a child is present' .
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected' .
FORMAT filter_$ (f1.0) .
FILTER BY filter_$ .
EXECUTE .
```

*When happy with the subpopulation, delete the unwanted cases .

*Syntax which deletes unwanted cases .

```
USE ALL.
SELECT IF( (relhrp > -1 & relhrp < 4 ) or (relhrp = 20) ).
EXECUTE .
```

Defining an indicator for the presence of a child

Ideally, we would use presence of an infant aged 0 (less than one year old). However, to avoid small number problems, define a variable which identifies the presence of a child age less than five years old. Have a look at the variables which indicate this.

Aggregating detailed ethnic group information

The ethnic group information in the LFS is detailed and with close equivalence to the 2001 Census. See below. The variable is called ethcen15. However, due to the sample size of the LFS there will be very small numbers in minority groups. Check this by a crosstabulation of ageq by ethcen15.

Value		Label
ethcen15	1	British
	2	Other White
	3	White and Black Caribbean
	4	White and Black African
	5	White and Asian
	6	Other Mixed
	7	Indian
	8	Pakistani
	9	Bangladeshi
	10	Other Asian
	11	Black Caribbean
	12	Black African
	13	Other Black
	14	Chinese
	15	Other

The following syntax aggregates ethcen15 into eight groups.

```
* 2001 broad ethnic groups .
compute ethgp = 8.
 if ( ethcen15 = 1 ) ethgp = 1 .
 if ( ethcen15 = 2 ) ethgp = 1 .
 if ( ethcen15 = 3 ) ethgp = 8 .
 if ( ethcen15 = 4 ) ethgp = 8 .
 if ( ethcen15 = 5 ) ethgp = 8 .
 if ( ethcen15 = 6 ) ethgp = 8 .
 if ( ethcen15 = 7 ) ethgp = 4 .
 if ( ethcen15 = 8 ) ethgp = 5 .
 if ( ethcen15 = 9 ) ethgp = 6 .
 if ( ethcen15 = 10 ) ethgp = 8 .
 if ( ethcen15 = 11 ) ethgp = 2 .
 if ( ethcen15 = 12 ) ethgp = 3 .
 if ( ethcen15 = 13 ) ethgp = 3 .
 if ( ethcen15 = 14 ) ethgp = 7 .
 if ( ethcen15 = 15 ) ethgp = 8 .

variable labels ethgp "Broad ethnic group" .
add value labels ethgp
        1 "White"
```

```
        2 "Black Caribbean"
        3 "Black African"
        4 "Indian"
        5 "Pakistani"
        6 "Bangladeshi"
        7 "Chinese"
        8 "Others".
execute.
```

Since even these groups may lead to small number problems, the following syntax creates broader groups:

```
* broader ethnic groups .
compute ethbgp = 4.
 if ( ethgp = 1 ) ethbgp = 1 .
 if ( ethgp = 2 or ethgp = 3 ) ethbgp = 2 .
 if ( ethgp > 3 and ethgp < 7 ) ethbgp = 3 .
 if ( ethgp = 7 or ethgp = 8 ) ethbgp = 4 .

variable labels ethbgp "Broader ethnic group" .
add value labels ethbgp
        1 "White"
        2 "Black"
        3 "South Asian"
        4 "Chinese & Others" .
execute.
```

OK, the data preparation is complete.

Some basic analyses to indicate fertility levels by ethnic group

Crosstabulations of ethnic group by presence of child can be used to inform a Child : Woman Ratio. This is likely here to be realistic by the eight ethnic groups defined above. Use these in conjunction with the national all women TFR to estimate TFRs by ethnic group. Compare these with the population-based TFRs by ethnic group and consider adjusting the latter since the survey directly links a woman's ethnicity with the presence of a child.

Crosstabulations of five year age by ethnic group by presence of child can be used as an age-specific fertility rate. These might be more reliable if calculated for the broadest ethnic groups. These can be the input to the Hadwiger mathematical function described below.

Some more advanced techniques to indicate fertility levels by ethnic group

The first section below is a statistical modelling approach, the second section uses a mathematical function.

1.) Statistical modelling: binary logistic regression

The probability of the presence of a child by ethnic group and age can be modelled

The approach is to select ethnic groups, one at a time by using a filter, for example, for the Black group:

```
*To select the Black ethnic group .
USE ALL.
COMPUTE filter_$=(ethbgp = 2).
VARIABLE LABEL filter_$ 'ethbgp = 1 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMAT filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE .
```

With the outcome variable 'child' (presence of a child aged 0-4), the syntax for a binary logistic regression

model is as follows:

```
LOGISTIC REGRESSION VARIABLES child
 /METHOD = ENTER ageq
 /CONTRAST (ageq)=Indicator(1)
 /PRINT = CI(95)
 /CRITERIA = PIN(.05) POUT(.10) ITERATE(20) CUT(.5) .
```

Use Excel for the next phase:

The regression coefficients, the log odds can be used to derive probabilities using:

z = constant + the log odds (for each age category)

The probabilities are calculated as $\exp(z) / (1 + \exp(z))$

Now consider the triangulation section

2.) Mathematical function: Hadwiger curve

Three files are provided to assist in the estimation of rates using the Hadwiger function

An Excel file, Raw-rates-to-smooth.xls, into which you can paste the cross-tab derived ASFRs by broad ethnic group. (The file also has graphed outputs of previous Hadwiger estimations so you can paste over the data to graph your estimate.)

An SPSS file, Hadwiger-fertility-input.sav, into which you can insert the data from the Excel file data preparation.

An SPSS syntax file, Hadwiger-syntax.sps, which you can use to estimate the ASFRs

Now consider the triangulation section

c.) Triangulation

Take the Age-Specific Fertility Rates as estimated using the LFS and using the information for each ethnic group and for all women, scale the national all women rates by the ratio of the ethnic group age-specific information to the all woman information. Then scale these broad ethnic group ASFRs by the CWR-derived TFRs for more detailed ethnic groups.

First do this at national level, then use the estimated ethnic group ASFRs to adjust the local authority all woman ASFRs. Also, adjust these by the CWR-derived TFRs.

Check the number of births this would represent and compare the total with the birth counts.