Using the ONS Longitudinal Study in epidemiological studies
What is the Longitudinal Study?

• a dynamic record linkage study with linked census records of 1% of the E&W population

• started in 1971, so shortly will comprise of records spanning up to 30 years for >600,000 individuals

• complemented by event data
Event data

- Live births and stillbirths to LS mothers
- Infant deaths to LS mothers
- Deaths of LS members and their spouses
- Embarkations
- all available to end of 2000
- Cancer registrations (to end 1997)
Person included from 1971 Census
Census data includes:

- age, sex, marital status
- family, household or communal establishment type
- housing, including tenure, rooms and amenities
- country of birth, and in 1971 parent's country of birth
- ethnicity (1991 & 2001)
- educational qualifications (only higher level qualifications 1971 & 1981; all levels 2001)
- economic activity
- occupation and social class
- migration and travel to work
- religion (2001)
- caregiving (2001)
- marriage and fertility history (1971)
Death registration data

*(up to date to end 2000)*

For deaths of LS members; spouses of LS members; stillbirths, neonatal & postneonatal deaths to infants of female LS members (some for male); deaths of children < 16 to be added.

- Cause of death, main & associated (up to 8)
- Certification type (1993_)
- Place of death, including type of communal establishment & length of stay (1993_)
- Occupation/employment/social class of LS member & husband/father
Cancer registration data

*(up to date to end 1997)*

- Age at registration
- Place of birth & usual residence
- Registry Centre
- Occupation / employment / social class of LS member & husband / father
- Basis of diagnosis
- Behaviour of neoplasm

**From 1993:**

- Age at diagnosis
- Marital status at registration
- Ethnic origin at registration
- Registered at screening indicator
- Tumour cell grading
- Treatment type
CeLSIUS  (Centre for Longitudinal Study Information and User Support)

Provides help to academic users of the LS
Funded by ESRC and based at LSHTM

www.celsius.lshtm.ac.uk
CeLSIUS and the ONS LS

• User support programme needed because, for confidentiality reasons, the data can only be accessed within a secure area at ONS

• Complexity of data set also means advice and training needed

• Only aggregated data can be supplied to researchers
Data restrictions

- prevent identification of individuals
- no individual-level data released
- limits on complexity of data:
  - categories, top coding
- care with geographical data
- data aggregated as frequency records
### Example of frequency record

<table>
<thead>
<tr>
<th>cases</th>
<th>sex</th>
<th>agegp</th>
<th>sclass</th>
<th>carstair</th>
<th>ltill9</th>
<th>emiss</th>
<th>4111110021851</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Research questions

• Cross-sectional (snapshot) at census point 1971, 1981 or 1991
• Change: investigating same people at two census points
• Duration: mortality, fertility, cancer incidence and survival
  – need event data
Computing

• used to be primarily SPSS with some SAS support

• introduction of Stata some years ago when survival analysis of LS data became popular

• Celsius team now work almost exclusively in Stata
Celsius

1981 - 1991

- Cancer registration
- Death

stset

stsplit

81 material

91 material

- Effectively get a person-years analysis
can collapse further to produce hazard rates by follow-up and categories of interest and produce a lifetable or can model

cases
deaths  pyrs period yearfup agegp mstatus housefam
1 0 1 1981-1985 0 45-49 Widowed Living alone
1 0 1 1981-1985 0 45-49 Widowed Lone parent+others
1 0 1 1981-1985 0 45-49 Widowed Lone parent+others
1 1 .19438 1981-1985 0 45-49 Married Couple+others
1 1 .64339 1981-1985 0 45-49 Married Couple+others
1 1 .78028 1981-1985 0 45-49 Married Couple+others
1 1 .02190 1981-1985 0 45-49 Married Couple no others
1 0 1 1981-1985 0 50-54 Single Living alone
1 0 1 1981-1985 0 50-54 Single No family but others
1 1 .01906 1981-1985 0 50-54 Single Communal estab.
1 0 1 1981-1985 0 50-54 Married Couple no others
1 0 1 1981-1985 0 50-54 Married Couple no others
<table>
<thead>
<tr>
<th>cases</th>
<th>deaths</th>
<th>pyrs</th>
<th>period</th>
<th>yearfup</th>
<th>agegp</th>
<th>expected mortality</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1981-1985</td>
<td>0</td>
<td>45-49</td>
<td>0.00034</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1981-1985</td>
<td>0</td>
<td>45-49</td>
<td>0.00034</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1981-1985</td>
<td>0</td>
<td>45-49</td>
<td>0.00034</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1981-1985</td>
<td>0</td>
<td>45-49</td>
<td>0.00034</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.19438</td>
<td>1981-1985</td>
<td>0</td>
<td>45-49</td>
<td>0.00034</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.64339</td>
<td>1981-1985</td>
<td>0</td>
<td>45-49</td>
<td>0.00034</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.78028</td>
<td>1981-1985</td>
<td>0</td>
<td>45-49</td>
<td>0.00034</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.02190</td>
<td>1981-1985</td>
<td>0</td>
<td>45-49</td>
<td>0.00034</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1981-1985</td>
<td>0</td>
<td>50-54</td>
<td>0.00072</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.01906</td>
<td>1981-1985</td>
<td>0</td>
<td>50-54</td>
<td>0.00072</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1981-1985</td>
<td>0</td>
<td>50-54</td>
<td>0.00072</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1981-1985</td>
<td>0</td>
<td>50-54</td>
<td>0.00072</td>
</tr>
</tbody>
</table>

can proceed to relative survival model

need to aggregate further
This data is OK to release and can form the basis of a Poisson model of crude survival or, by using a user defined link function in a glm model, relative survival can be modelled (year of follow up always included in model)


Accepted for publication 2003
Table 1: Crude and relative survival at 1, 5 and 10 years: comparison with Cancer Survival Trends (CST)* data for breast cancer.

<table>
<thead>
<tr>
<th>Survival Time</th>
<th>Crude survival (%)</th>
<th>Relative survival (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>This study 1981-1997 (95% CIs)</td>
<td>CST 1981-85</td>
</tr>
<tr>
<td>1 year</td>
<td>86.4 (85.5-87.2)</td>
<td>85</td>
</tr>
<tr>
<td>5 year</td>
<td>58.4 (57.0-59.8)</td>
<td>54</td>
</tr>
<tr>
<td>10 year</td>
<td>42.4 (40.2-44.6)</td>
<td>37</td>
</tr>
</tbody>
</table>

* (Coleman et al. 1999)
Using the LS

• Decide if the LS is a suitable source for the research project planned:
  consult website; review publications (online list);
  consult CeLSIUS staff

• If yes, prepare project outline, consult CeLSIUS staff and submit proposal, including confidentiality undertaking

• Specify data extract needed using online resources (data dictionary) and help from assigned CeLSIUS support officer

• Specify analyses needed/aggregate tabulations

• Notify CeLSIUS of publications and presentations