Transforming Everyday Life into Extraordinary Ideas
Sampling and Recruitment in the CLSA

Lauren Griffith, Harry Shannon, Parminder Raina, Christina Wolfson, Susan Kirkland

Gerontological Society of America (GSA)
San Diego, Nov. 15, 2012
Relevance

• Sampling underlies our ability to provide reliable estimates using the CLSA data
• We will need to use weights to estimate parameters (means, proportions, etc) for the target population
• Weights involve:
  • Calculating probability of selecting sampling unit (in CLSA, unit = person)
  • Accounting for different sampling frames
  • Allowing for non-response
Aims of sampling in CLSA

- Choose representative sample of eligible Canadians
  - 20K Tracking cohort; 30K Comprehensive cohort
  - Specified numbers in age-sex groups by province
Potential Sampling Frames

- Canadian Community Health Survey Participants
- Provincial Health Registration Databases
- Random Digit Dialling

ALL OF THE ABOVE
# Canadian Longitudinal Study on Aging

<table>
<thead>
<tr>
<th>Sampling Frame: CCHS, provincial health registration databases, and RDD</th>
<th>Sampling Frame: provincial health registration databases, and RDD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CLSA Tracking (n=20,000)</strong></td>
<td><strong>CLSA Comprehensive (n=30,000)</strong></td>
</tr>
<tr>
<td>45-54</td>
<td>45-54</td>
</tr>
<tr>
<td>55-64</td>
<td>55-64</td>
</tr>
<tr>
<td>65-74</td>
<td>65-74</td>
</tr>
<tr>
<td>75-85</td>
<td>75-85</td>
</tr>
<tr>
<td>↓</td>
<td>↓</td>
</tr>
<tr>
<td>6,000</td>
<td>9,000</td>
</tr>
<tr>
<td>6,000</td>
<td>9,000</td>
</tr>
<tr>
<td>4,000</td>
<td>6,000</td>
</tr>
<tr>
<td>4,000</td>
<td>6,000</td>
</tr>
</tbody>
</table>
CLSA Sampling Frames

• CCHS provided first part of sample
• Options for methods of selection of remaining participants:
  • Using provincial health registries - preferred
  • Random digit dialing
• In several provinces, we cannot use registries, so need to do RDD
Recruitment from the CCHS

- CLSA collaborated with Statistics Canada to develop the CCHS Healthy Aging Questionnaire

- **Target population**: People aged 45 and over living in private occupied dwellings in the ten provinces

- **Excluded**:
  - Residents of the three territories
  - Persons living on Indian reserves or Crown lands
  - Persons living in institutions
  - Full-time members of the Canadian Forces
  - Residents of some remote regions
Recruitment from the CCHS, *ctd.*

Participants were asked to share:
- Their contact information with the CLSA (for recruitment)
- Their survey responses with the CLSA (for analysis)

- N=30,865
- N=26,248
- N=4,617
- >85
- N=11,742
- N=527
- N=8,345
- N=5,634
Recruitment from the CCHS, *ctd.*

### Canadian Longitudinal Study on Aging

| Sampling Frame: CCHS, provincial health registration databases, and RDD |
|---|---|---|---|---|
| CLSA Tracking (n=20,000) | 45-54 | 55-64 | 65-74 | 75-85 |
| CCHS | 617 | 1,704 | 1,350 | 791 |
| Remainder | 5,383 | 4,296 | 2,650 | 3,209 |

| Sampling Frame: provincial health registration databases, and RDD |
|---|---|---|---|---|
| CLSA Comprehensive (n=30,000) | 45-54 | 55-64 | 65-74 | 75-85 |
| | 9,000 | 9,000 | 6,000 | 6,000 |
Recruitment from Provincial health registration databases

- **2005**
  - Feasibility study to explore practical, methodological and ethical aspects of accessing Health Care Utilization data from Provincial databases (published 2009)

- **2009-2011**
  - Several meetings with Provincial Data Stewards and Privacy Commissioners to negotiate access to health registration databases for sampling
MOH Mailout
• Letter(s) from MOH and/or CLSA
• Brochure, Information Package
• Consent to contact form
  + Postcard reminder in 20 days

Participant returns consent to contact form

NCC
• Assigns unique ID
• Sends:
  • Participant consent form
  • Additional study materials

Participant contacted

CATI
• Assess eligibility
• Answers participant questions

Participant interested and ready

Participant interested but not ready

CATI
• Participant consent
• Conducts telephone interview
• Collects Provincial Health Number
  (if participant provides consent)

CATI
• Schedule Interview
RDD approach

• In principle, idea is simple
• Randomly sample numbers as far as possible in specified area codes and with next 3 digits in relevant area
• Identify eligible people at each number
• Randomly choose one person
• Recruit willing participants until ‘quota’ filled
Original Plan for Additional Recruitment

P=Provincial Health Registration Databases
R=RDD Only
H=Hybrid: RDD then Provincial Health Registration Databases
<table>
<thead>
<tr>
<th>Cohort</th>
<th>Recruitment Method</th>
<th>Initial Response Rate Range</th>
<th>Mailout</th>
<th>Recruitment Rate</th>
<th>Mailout</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tracking</td>
<td>MOH</td>
<td>LOW 1.7%</td>
<td>ON July 2012</td>
<td>LOW 45.5%</td>
<td>NL Apr 2012</td>
</tr>
<tr>
<td>NB, PE, MB, NS, NL, SK, ON</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comprehensive</td>
<td>MOH</td>
<td>LOW 4.4%</td>
<td>MB May, Aug 2012</td>
<td>LOW 29.3%</td>
<td>NL Apr, July 2012</td>
</tr>
<tr>
<td>NS, NL, MB, ON</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tracking</td>
<td>RDD</td>
<td></td>
<td></td>
<td>LOW 33.6%</td>
<td>ON Apr 2012</td>
</tr>
<tr>
<td>AB, BC, QC, ON</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>QC Mar 2012</td>
</tr>
<tr>
<td>Comprehensive</td>
<td>RDD</td>
<td></td>
<td></td>
<td>LOW 35.4%</td>
<td>AB Mar 2012</td>
</tr>
<tr>
<td>AB, BC, QC, ON</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ON Apr 2012</td>
</tr>
<tr>
<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>
</tr>
</tbody>
</table>
Combining samples from different sources

- Want overall $P(\text{Participation})$
- Use addition rule of probability
- E.g., for someone chosen via RDD, need $P(\text{Selected by RDD}) \text{ AND } P(\text{Selected in CCHS})$
- Latter is an average probability, not an individual one
- Similarly for selection through HR
Issues

- When $P(\text{Participation})$ is based on the product of probabilities, have to assume independence of probabilities
- Confidentiality conditions may mean, e.g., we call people in RDD who were in the CCHS and did not want to participate in the CLSA
- In RDD, have to allow for multiple phones in the household
- As in many population-based studies, recruitment in lower SES categories is difficult
Summary

• Various sources of participants for CLSA
• Each has its own strengths and limitations
• Need to estimate sampling probabilities for each source
• Aiming for representativeness – but …
• Various assumptions must be made
Transforming Everyday Life into Extraordinary Ideas

www.clsae-lcve.ca