The Quality of Survey Data

Mary Thompson

January 11, 2006

Importance of observation

- Evidence-based practice requires data
- Data come in many forms
- Observational studies supplement experimental studies

EXAMPLE: Ephedra

- Morton, S. C. (2005) *Statistical Science*
- Dietary supplement
- Clinical data inadequate to assess safety
- Need for "observational data" (adverse events) from the population

Surveys …..

- Provide observation on large populations
- Tell us to what extent the phenomena we observe are replicable: "the syme the whole world over" -- W. G. Cochran

What constitutes a survey?

Survey elements:

- Study population
- Sample chosen through probability sampling design
- Efforts made to measure characteristics of sample members
- Generalization to population of interest

What is Meant by "Quality"?

We consider some examples of survey quality ….

Example: Prevalence of herpes

Six cities survey:

- 6 PCP randomly selected offices in relatively affluent areas of each of Atlanta, Baltimore, Boston, Chicago, Dallas, Denver

- Volunteers recruited, 150 per PCP

- 5732 screened, 5433 in final sample

Example: Prevalence of herpes

- The overall weighted HSV-2 seroprevalence was 25.5 percent
- 95% confidence interval 20.2 per cent to 30.8 per cent

Example: Prevalence of herpes

NHANES III:

- probability sample of 40,000 people across the U.S. (first stage units are counties, about 80)

- Blood tests (response rate 75%)

- Estimated 21.3% of U.S. population over 12 were positive for HSV-2 in 1991

- 95% CI 19.4% - 23.3 %

Example: IDUs

- Needle exchanges, community drug teams, drop-in centres, out-patient clinics

- Survey on safer injection facilities in Vancouver 2003-2004:   about 450 IDUs recruited by street recruitment and "snowballing"

Example: Alcohol problems

- Survey to measure prevalence of problem drinking?

- NHANES Mobile Examination Center

- NHANES IV 1999-2000

- CCHS 1.2 mental health and well-being: 34 questions

Example: Smoking behaviour

Prevalence of daily cigarette smoking in Canada, adults:

- ITC: RDD, 50% response rate
- CTUMS: RDD, 77.5% response rate
- CCHS: mixed, 85% response rate

Example: Physical activity

- CCHS 1.1   80% area frame, 20% RDD; 50% CATI

- CCHS 2.1   50% area frame, 50 % RDD; 70% CATI

(Mode effect)

Components of survey error

- Coverage error   (six cities)
- Non-response error (RDD survey)
- Measurement error (activity)

Components of survey error

Sampling error:

- … precise "within 3 percentage points, 19 times in 20" (opinion polls)

- Clustering of sample tends to increase sampling error (six cities)

Components of survey error

"Gremlins" or human error:

- Choosing units
- Data collection
- Data recording
- Data storing and retrieval
- ….

FIFTH CENSUS OF CANADA, 1911.

SCHEDULE No. 1
TABLEAU

CINQUIÈME
DU CAN...

Population by Name, Personal Description, etc.
Population—Nom, Renseignements Personnels, etc.

Province Ontario   District No. 7 3 Haldimand   Sub District No.   Enumeration District No. 2   in / dans Nanticoke Town   (City, town, village, township or parish.)

Page 4

Nominal return of living persons by / Dénombrement des vivants par : Thomas Boone, Enumerator.

No.	No.	Name of each person	Place of habitation	Rel.	Relationship	Sex	Month of birth	Year of birth	Age	Country or place of birth	Cit.	Year of immigration	Year of naturalization	Racial or tribal origin	Nationality	Religion	Occupation	Employment	Earnings	Col 19	Col 20	Col 21	Wage earners col 22	Col 23	Col 24	Col 25	Col 26	Col 27	Col 28	Col 29	Insurance col 30	Col 31	Col 32	Col 33	Col 34	Col 35	Col 36	Col 37	Col 38	Col 39
1	1	Ross Sarah				F	Oct	1859	51	O				Irish		Methodist	Dressmaker	9-0-886	yes	no	yes												yes	yes		
2		" Annie				F	July	1889	21	O				Irish		Methodist	none	nil	no	no													yes	yes		
3	2	Gibson Susanna			Head	F	Feb	1840	70	England 1872				English		Anglican	none	nil	no	no													yes	yes		
4		" Walter			Son	M	Sep	1864	27	O				English		Anglican	Magician	16-30	yes	yes					100		200						yes	yes		
5		" Maud			Daughter	F	Feb	1868	23	O				English		Anglican	none	nil	no	no													yes	yes		
6		" Alfred			Son	M	Dec	1887	22	O				English		Anglican	Barber	10-6-18	yes	no					100		400						yes	yes		
7	3	McNally Mary			Head	F	Mar	1855	51	O				Irish				nil	no	no													yes	yes		
8		" Mary			Daughter	F	Oct	1882		O				Irish				10-7-77				Gen Store		232									yes	yes		
9		" Etheldreda			Son	M	Nov	1876		O				Irish				nil	no	no													yes	yes		
10	4	Ross Thomas			Head	M	Mar	1835		O								1-9-00	yes	no	yes	Wood Labor		100									yes	yes		
11	5	Eastland Thomas			Head	M	Feb	1839		O				English				18-0-77				Gen Store			2000		53.60			yes	yes		
12		" Maud			Wife	F	Jan	1852		O				English				10-0-77	no	no		Gen Store								yes	yes		
13		" Bell			Sister	F	Feb	1852	59	O				English			Merchant	10-0-77	no	no		Gen Store								yes	yes		
14	6	Lovell Coleman			Head	M	Dec	1861	49	O							Merchant	10-0-	yes	no		Jeweler				2000		105			yes		

Assessment of survey quality?

- Do the conclusions stand up under scrutiny?
- Could a hostile critic explain them away?

Achieving Survey Quality

Classical: examine the components of estimation error

- measurement error

- sampling error

- non-response error

- coverage error

- gremlins

Try to minimize each one

Reference

- Paul P. Biemer, Lars E. Lyberg: Introduction to Survey Quality

- TQM approach

- Typical goals: improve staff training; streamline questionnaires; introduce more checks on data entry

Example: ITC China survey

- Detailed instructions ("please write legibly")
- Carefully piloted data collection materials
- "Hands-on" training at all levels
- Audio recording of interviews
- ….

Total survey error approach

Weisberg, H. F. (2005) University of Chicago Press:

- " … A guide to the new science of survey research"

- Survey errors, survey constraints (cost, time, ethical considerations), survey related effects

Example: RRFSS

Northrup et al (2004) Response rates and results of the advance letter experiment

- "What did it take to get the 62% response rate for the RRFSS?"

Statistical support

The role of statistics:

- Quantify survey error

- To the extent possible, "correct" inferences taking into account missing data and measurement error

Example: Latent class analysis

Biemer et al (2002 JRSSA; 2005):

- Estimate prevalence of drug use by using several "flawed" measurements, e.g. direct questions and randomized response questions

New Journal

Journal of Obnoxious Statistics (2005):

- "Perfect quality in the Swiss deep sea fishing survey"

- "Non-response in surveys of the dead"

Quantifying survey error

- Probability sampling is important
- Minimizing non-sampling error is important
- Quantifying non-sampling error requires careful modelling, sensitivity analyses
- Quantifying total survey error requires full knowledge of the design, the operations, the editing and imputation process

Example: Satisfaction survey

- Mazor et al (2002) *Health Services Research*

- Analysis of patient satisfaction data, modelling of response probabilities, estimation of non-response bias

- Estimates of satisfaction may be most inflated for providers with the least satisfied patients .....

In the end …..

- The quality of survey data depends on the purposes and the conclusions, as well as the methods and operations.

- The "science" of surveys is coming into its own.