## 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
- ....

					STATE OF THE PARTY			William Street	deposit									0			IN A STREET OF THE STREET			SO PARTY.			SELECTIVE OF THE PARTY OF THE P		MALVAN A
FIFTH CENSUS OF CANADA, 1911.			SCHEDULE NG 1. POPULATION ST NAME, PERSONAL DESCRIPTION, ETC. TABLEAU NG 1. POPULATION ST NAME, PERSONAL DESCRIPTION, ETC. TABLEAU NG 1. POPULATION-HOM, RENSEIGNEMENTS PERSONALS, ETC. Formalian District Name (Sales) Na (											un.	CINQUIÈN  ((Vey, 100 m., village, toreschip or pichla.)  ((City, 100 m., village, canaca on paraisso.)														
· ·	e 4	Provi	DENCE AND PLES	10	1	174trii				S-FLG.V.				Distri	Enur Enur Roots		FLIVING	į dans į	WAGI	EABNERS				SELD AT DATE			EARS OF AU		ape.)
	CHIRALIAN DESCRIPTION OF THE PROPERTY OF THE P	Name of such power in family, books shall in artifaction.	Phoese of Buddenties, Community or partials, stiff, Emigra or reseasement, and Silver contenties, between the contenties, and the contenties of the contenti	Bes. Se bend of banks	housing or the state of the sta	Menth of birth.	Treated Meth.	Character of pines of lists. dl in Catacht specify operation as negatively.	Trace of interpretation to the control of the contr	Rapid or tribo origin	Nation	eliky.	Brigin.	State equation of the	Sections of an	ent other to the total to	Supply to the sound of the soun	Rays where person by exploring as to have "in resolve gall," and herefore stage, "in seven	Weeks conjured to the action of the consparing of the consparing of the consparing of the constant of the cons	Heart Accessing the formal per water at Accessing the accessing the accessing the accessing the formal accessing the formal accessing the formal accessing the accessing the formal accessing the formal accessing the formal accessing the formal accessing to the formal accessing the formal accessing the formal accessing the formal accessing to	portgaine, V von Totaleneline is 300 Totaleneline is 300 Post central or retain	Second manager for particular and particular for particular and particular for pa	EANCE EX	VIGUEUR A D.	Notice at entropy in	TOOLESS !	Tangers or the state of the sta	Complete Committee Committ	w INP
1	THERESTER IN LOST OF THE PROPERTY OF THE PROPE	New Or disagns personne does in tracing.  In the tracing on Emeritation.	Life on Demicies, GDA, Cartini on participation, GDA, COR on Village, COR ON VILLAG	Sean. Proventif a di particolo del monto.	Security of the second security of the second security of the second sec	Esta de naimeaca.	Atrice for responses.	Prope on Dieg de Salestive, (Se lieu de rader rere en; le vanada, ree- cifar la pessition du le territaire).	Section for the production of	Origina salan ia ra-		azite.	tietges.	Autoritacipa es re	Employ as Easter on the Other Femploy per dis not will y a	politiment di information di di interipal di di di citer, di	Travillant a ten remotes (vetr instructions)	Mentioner of he permission of the permission of capital de capital	Nombre de amplhos Sommer en misson 1 inspector de amplhos 1 inspector de amplhos Sonder de amplhos Antonios compais	Moure do provide a provide	A function of the state of the	Charleton per lever Spark de policie per lever Mentre e Cerritas	-	at the south of th	a Sandara de maria	34 3	had after the same and same an	FALLE Memorina or 100 per de persona e 100 persona	Table of the same
-	2 2	Herr Faran n Munici Gelson temma n Halter n Mand	18	2 160 16 300	ship of		1859 5 2 1897 1 2 847 6 2 584 2 1 88 23	Congland &		Joseph Brightsho English English	Cana	dien leng dienkang	thebist lica ve lica ve lica ve	Jane Jane	1-0- ng	-36 n	no n				2.00				۶	300 3 300 3 300 3	40 E		
8 . 9 . 10	4 4	the Alfred War War of W	44	the de		1665 1 1665 1 1660	88625	(s) (s) (s) (s)		English	leans Land	do sea dans	g butoria. Animana Marina.	Land Confe	1-9	-77 no	no mo	son Mise Odd fol-	1		222 158	Z	Care.	3 x	-	300 3 300 30 300 30 300 30			
12 13 14	6 6	" Short Provide Bayes	is. Vendia	1 2 3 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	lisa Ja		7553 Z.5 1592 79 7561 49	9 9 9		Errore	Charles	dear the	esta mon	Marchan	100	77 20	na year	Sin Gloric Lunder					ere .	10		300 p	4 &		
										2044A	Chicago Garage Garage Garage					4/ 0	4												
				* *				1																					
						,										asije s			ME S									2 31	
	T 14								is Si					100															
															e ele									05 to					
							e de la companya de l			7.5	Sec. 25 40 2																*		
																	A-10-	<b>,</b>											

### 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.