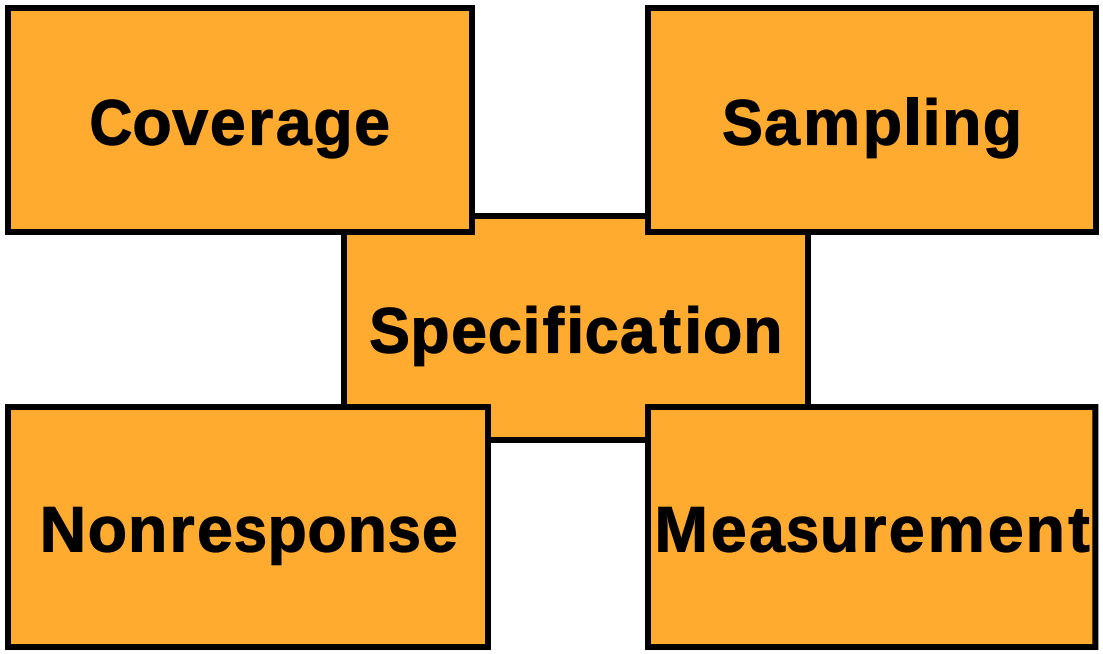
The Cornerstones of Data Quality
Do's and Don'ts In Survey Research

Edith D. de Leeuw

Share and Discover Seminar Team Vier

9 December 2008

Cornerstones Survey Data Quality



Importance Cornerstones

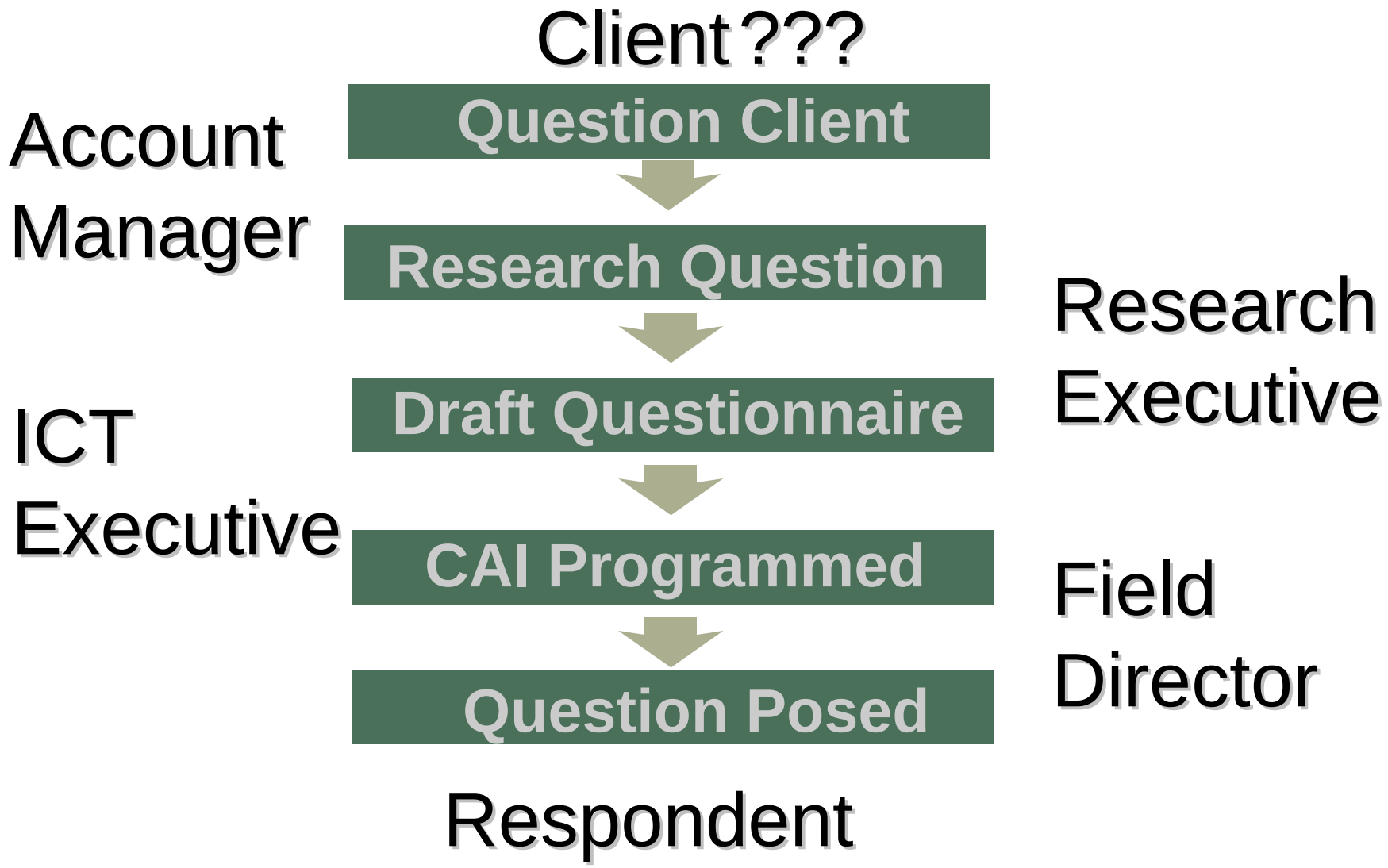
- Potential Sources of Error
  - Threat to Survey Quality
- Analyze at start research project
  - How important is this specific cornerstone?
  - Is potential threat large?
  - What can I do to control this source of error ?
    - AS BEST AS POSSIBLE
- Consider costs and errors!
  - See, Groves, 1989

Specification

- Foundation is specification
- Specification error occurs when the concept measured by a survey questions and the concept that should be measured with this question differ
  - You ask the wrong questions!
  - Validity problem
  - Your client gets an answer to a question s/he did not ask

De Leeuw, Hox, Dillman, 2008

Specification needs Team Work and Communication



Coverage

- Coverage
  - The percentage of the population of interest that is included in the sampling frame
- Coverage Error
  - Sampling frame must include *all* units of population of interest (once), if not coverage error
    - Undercoverage
    - Overcoverage

De Leeuw, Hox, Dillman, 2008

Coverage 2

- General population may have problems:
  - Institutionalized
  - No web access
  - Mobile only, etc
- Analyze if problem
  - People in old-age homes and consumption electric appliances?
    - Not really
  - Elderly and consumption of medical care?
    - Probably
- Potential solution
  - Mixed-mode, dual frame sampling!

Coverage 3

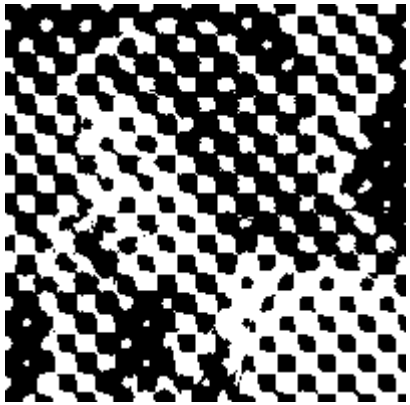
- Special populations
  - Well defined lists, but check!
- Client provides sampling frame
  - Age of information
  - Duplications
  - Erroneous inclusions
- Check!

Sampling

- Sampling Error
  - Occurs because only a sample of the population is invited to participate in a survey in stead of the total population:
    - Statistic of interest is computed on sample
- Probability Sampling
  - Standard error, confidence intervals
- Non probability Sampling
  - No statistics, no standard errors

De Leeuw, Hox, Dillman, 2008

Probability Sampling
a visual representation



Source: Dr Andrew Balemi, http://www.stat.auckland.ac.nz/~balemi/)

A Large N

- Sample size and Sampling Error
  - Larger N means smaller sampling error
    - SE is standard deviation divided by square root of N!
  - Clearer picture
  - Probability sampling only!

- “100.000 people can not be wrong”
  - Yes they can!
    - If (self) selective nonprobability sample

Non Probability Sampling

- Any of several different sampling schemes in which the elements in the "sampling frame" do not have both a known and and a nonzero probability of selection.
- Thus it is impossible to calculate the size of a poll's margin of sampling error with a nonprobability sample. Of note, this statistical fact does not stop some pollsters from calculating sampling error with a nonprobability sample - it just makes their calculations meaningless.
- Nonprobability samples are useful in the early stages of research or when a pollster needs to gain an "impression" of the preferences and attitudes of a target population but does not need to be very confident about how well the poll generalizes to the target population.

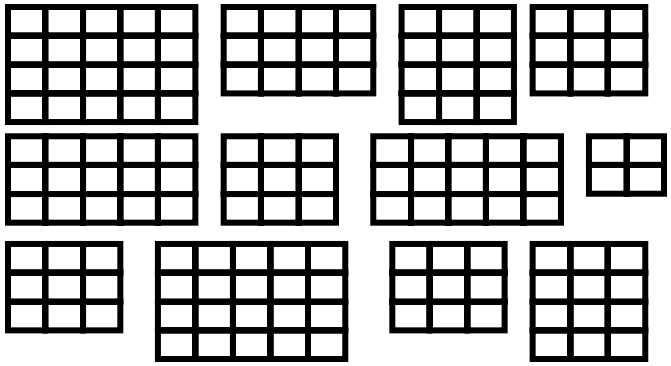
www.aapor.org, glossary dec 2008

Sampling and Efficiency

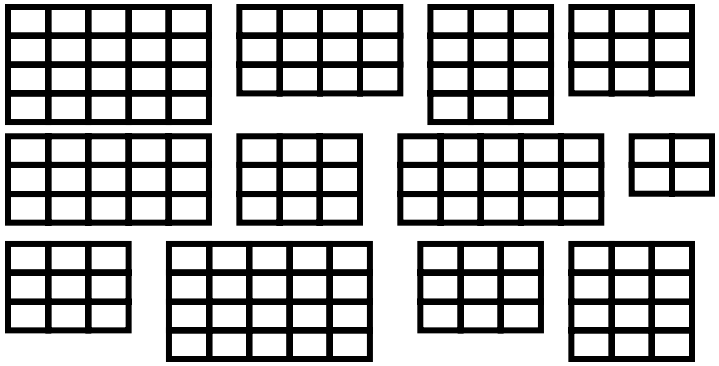
- Simple Random Sampling (SRS)
  - Easy, but sometimes not efficient
- Cluster sampling
  - Efficiency in time
- Stratified sampling
  - Efficiency in sampling size
- Analysis needs to take sampling scheme inconsideration!

Source: Dr Andrew Balemi, http://www.stat.auckland.ac.nz/~balemi/)

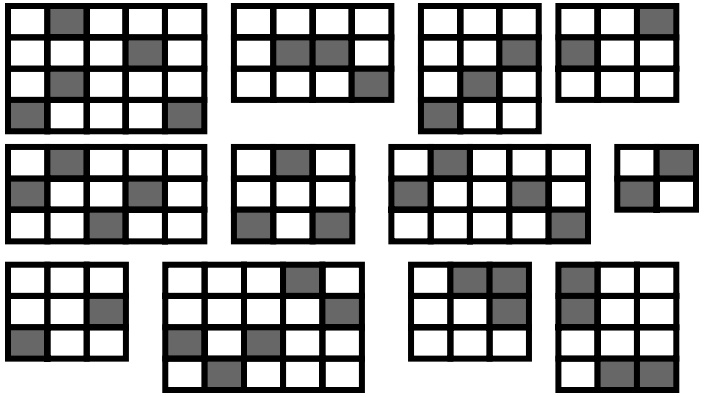
Stratified Sampling Cluster Sampling



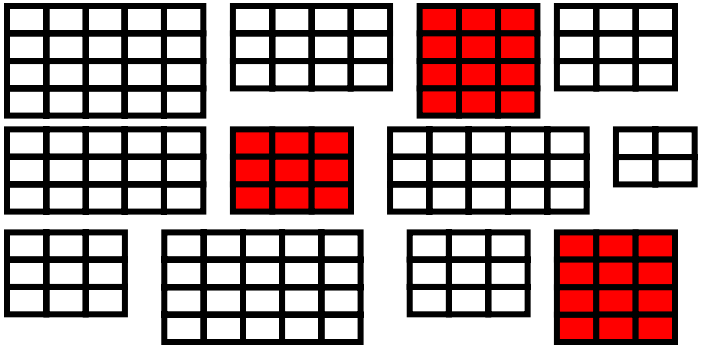
Population of L strata, stratum l contains n_l units



Population of C clusters



Take simple random sample in *every* stratum



Take srs of clusters, sample every unit in chosen clusters

Report Nonresponse

- Response Rate Calculation
  - General overview AAPOR at www.aapor.org
  - Response rate web surveys: ESOMAR
    - http://www.esomar.org/index.php/professional-standards.html
      - Conducting research using the internet
    - See also EFAMRO quality standards
      - http://www.esomar.org/index.php/professional-standards.html
- Report different sources of (Non)Response
  - Not located/wrong address/bounced
  - Not reached, not at home
  - Refusal, etc

Nonresponse and Error

- Nonresponse Error
  1. Nonresponse occurs
  2. Respondents and non-respondents differ on variable of interest (key variable study)
- In other words
  - Nonresponse error is the difference between a survey estimate and the equivalent estimate that would have been obtained if all selected units had responded

De Leeuw, Hox, Dillman, 2008

Nonresponse Studies

- Investigate nonresponse
  - Compare on background characteristics
  - Information in sampling frame
  - Eager vs reluctant respondents
    - Early vs late respondents
    - Immediate vs conversed respondents
  - Survey subsample nonrespondent

Adjustment

- Adjust for nonresponse
  - Choose weighting/ adjustment variables well
- Correction variable should
  1. Predict nonresponse
  2. Predict outcome variable of interest
     - If not related to outcome variable
       - Only cosmetic, looks representative on background variables, but……

Measurement

- Measurement Error:
  - Lack of reliability and validity

- Potential Sources of Measurement Error
  - Questionnaire
  - Respondent
  - Mode
  - Interactions, e.g., respondent questionnaire

De Leeuw, Hox, Dillman, 2008

Questionnaire

- Carefully constructed and tested questionnaire best prevention!
- Write simple questions, in simple words for simple people
  - Even for higher educated respondents ☺

Pretest Questionnaire

Even after years of experience, no expert can write a perfect questionnaire…..

Sudman & Bradburn , 1982

Testing is the only way of assuring that the survey questions written, do indeed communicate to respondents as intended

Campanelli, 2008

Pretest Methods

- Many methods available
- Each has different goal/purpose
- At different time during questionnaire production
  - Informal Methods
  - Pretest
    - Expert reviews
    - Cognitive interviews
  - Functionality check
  - Use-ability test
  - Pilot or Field test

Document & Communicate

- Total Quality Management
  - Documentation
  - Communication
- Documentation
  - Quality Handbook
  - Registration of complaints, customer satisfaction
  - Research report
- Communication
  - Courses
  - Meetings
  - Seminars

Further Readings

- Edith De Leeuw, Joop Hox & Don Dillman (2008)
  - International Handbook of survey Methodology.
    - New York: Lawrence Erlbaum/Psychology Press, Taylor and Francis Group
- Don Dillman, Jolene Smyth, & Leah Christian (2009)
  - Internet, Mail, and Mixed Mode Surveys
    - New York: John Wiley & Sons

Helpful Websites

- Homepage Edith de Leeuw
  - http://www.xs4all.nl/~edithl/
    - Summary International Handbook of survey methodology: introduction, chapter summary, glossary, additional material
- Don Dillman website available papers
  - http://survey.sesrc.wsu.edu/dillman/
- Journal of Official Statistics (JOS)
  - http://www.jos.nu