



# **SURVEY DATA QUALITY CONTROL STRATEGIES: AN INSTITUTIONAL RESEARCH PERSPECTIVE**

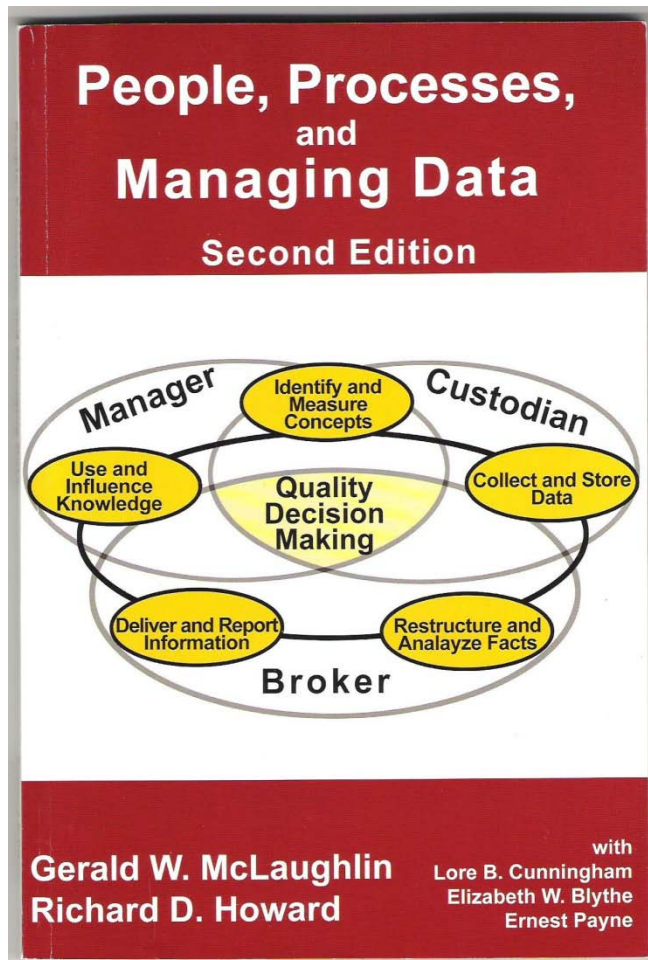
**Qin Liu**

**Research Analyst, Institutional Planning & Analysis Office  
British Columbia Institute of Technology, Vancouver, Canada**

**1**

**Presented to the AIR annual forum  
Atlanta, June 3, 2009**

# BACKGROUND



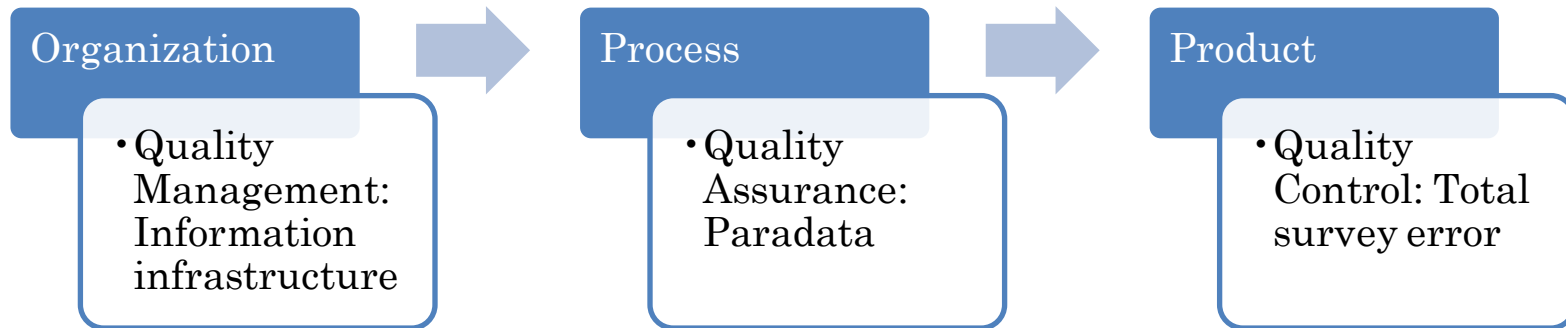
- Administrative vs. survey data
- Lack of synthesis

*Improve survey data quality!*

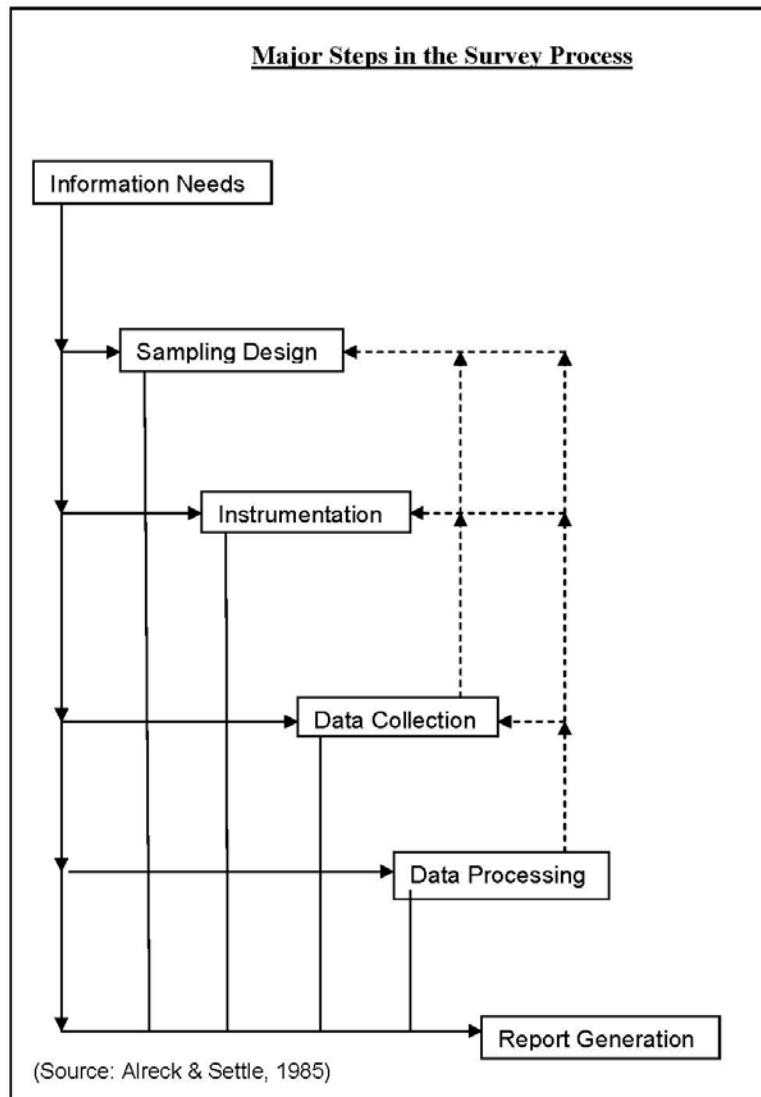
# Agenda

- Three levels of survey quality
- Survey process: two perspectives
- Total survey error
- Survey data quality strategy: quality assurance and quality control
- Data quality documentation
- An example
- Concluding remarks

# SURVEY QUALITY



# SURVEY PROCESS: TWO PERSPECTIVES



- Design perspective

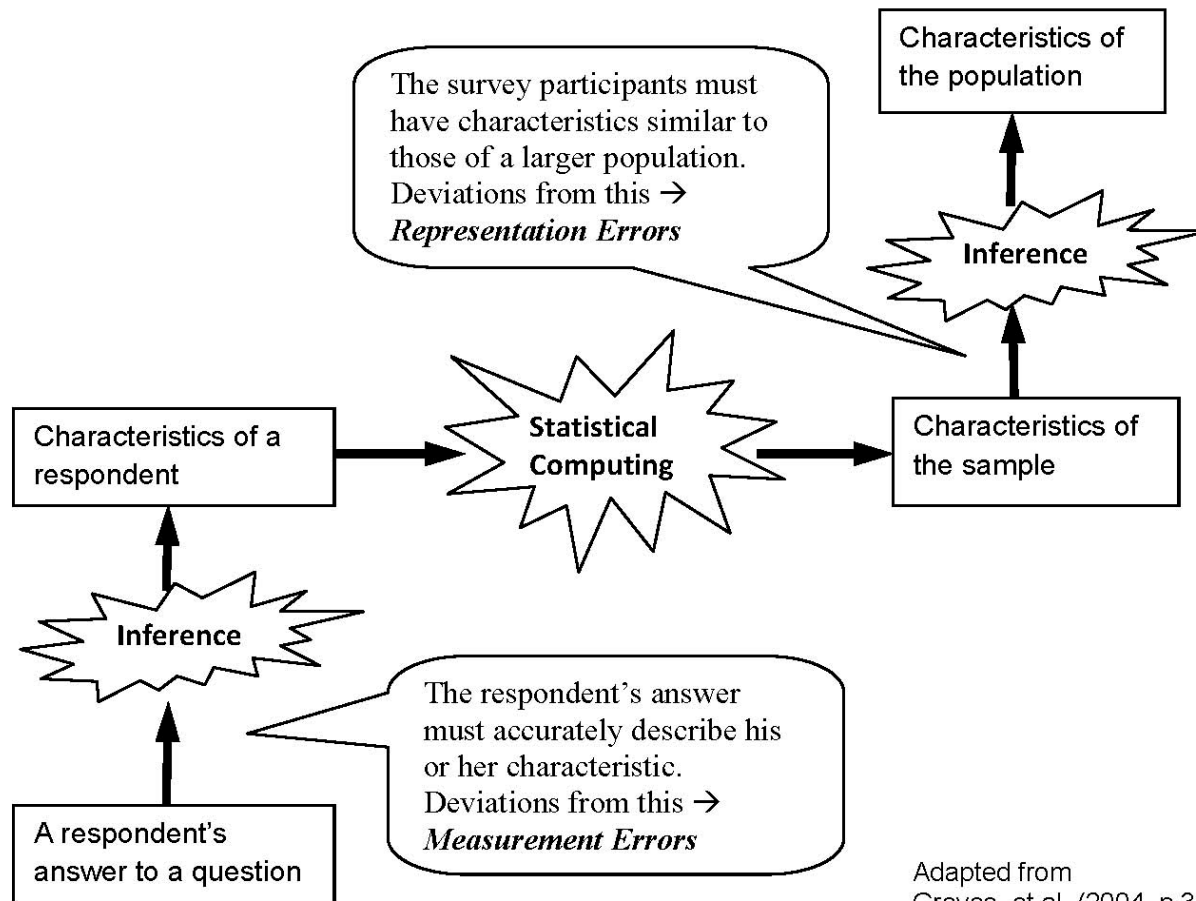
- Quality perspective

“Error”:

*Deviations of what is obtained in the survey process from what is desired*

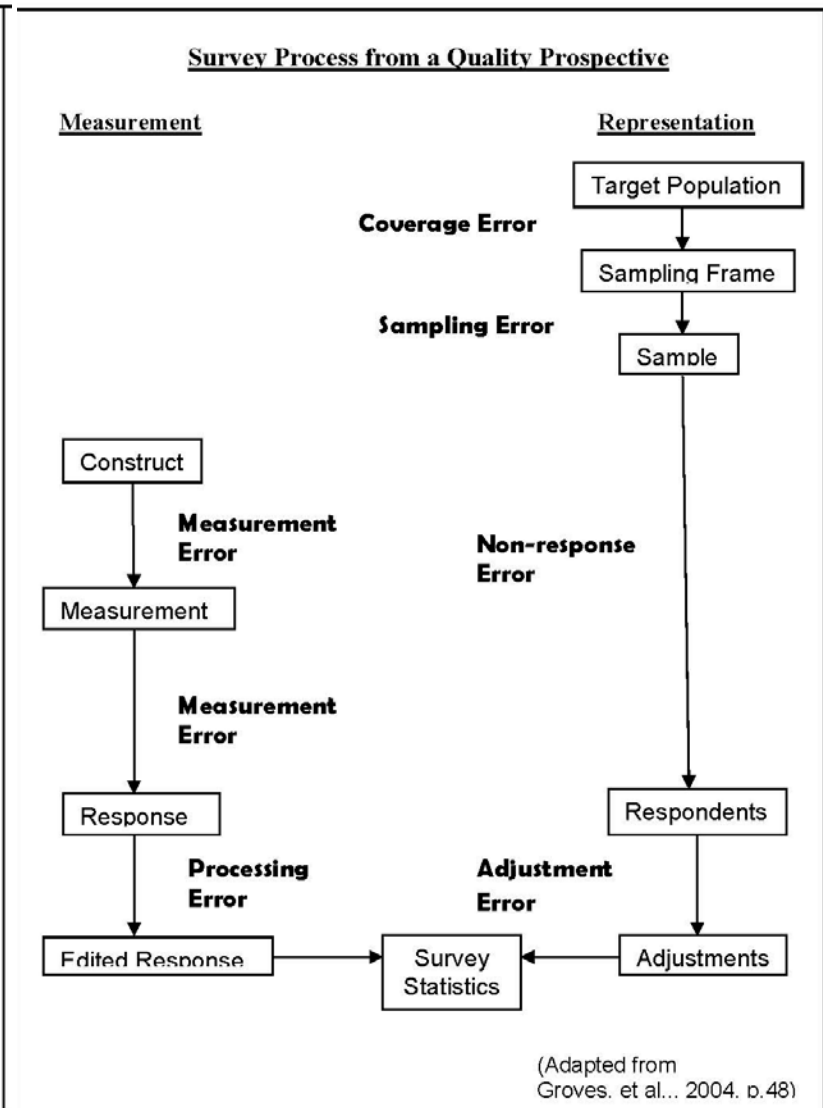
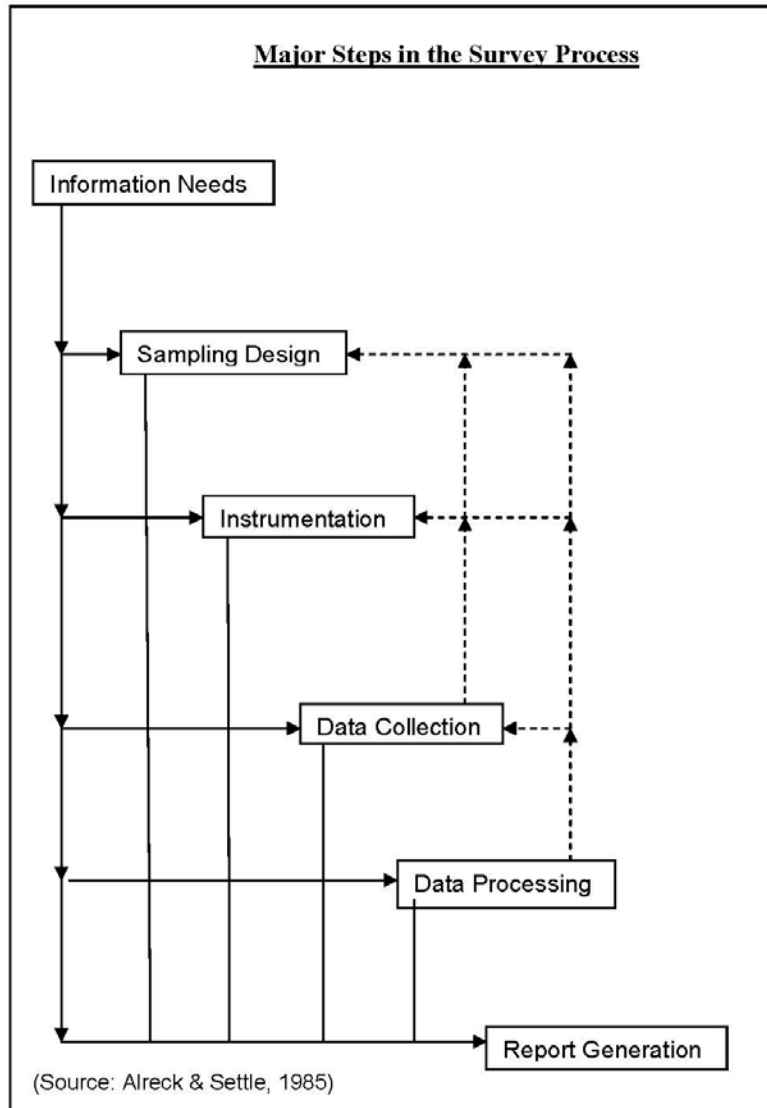
# TWO TYPES OF SURVEY INFERENCE

## Two types of survey inference



Adapted from  
Groves, et al. (2004, p.39)

# SURVEY PROCESS AND SURVEY ERRORS



# TOTAL SURVEY ERROR

- The task of a survey researcher:

to minimize error in survey statistics by making design and estimation choices that minimize the gap between two successive stages of the survey process

- “Error”:

Deviations of what is attained in the survey process from what is desired

# A SURVEY DATA QUALITY STRATEGY

- Measurement error
- Coverage error
- Sampling error
- Non-response error
- Processing/adjustment error (post-survey)

- Indicators of Quality Data
- Quality Control Procedures
- Quality Assurance Procedures

# To reduce measurement error

Indicators of Quality Data	Quality Control Procedure	Quality Assurance Procedures
<p>1). High validity</p> <p>2). Low response bias</p> <p>3). High reliability</p>	<p>Assess validity by checking Construct validity, concurrent validity, &amp; divergent validity;</p> <p>Assess response bias by</p> <ul style="list-style-type: none"> <li>• comparing survey data with data or information from sources external to the survey;</li> <li>• checking response tendencies</li> </ul> <p>Assess reliability by checking an index of consistency (e.g., Cronbach's alpha for internal consistency)</p>	<ul style="list-style-type: none"> <li>• Improve question wording;</li> <li>• Good questionnaire construction;</li> <li>• Conduct cognitive interviews;</li> <li>• Ensure adequate respondent behavior;</li> <li>• Ensure adequate interviewer behavior (if interviewer-administered)</li> </ul>

# To reduce coverage error

Indicators of Quality Data	Quality Control Procedure	Quality Assurance Procedures
The sampling frame is as close as possible to target population	Compare the specifications of the target population and the corresponding parameters of the sampling frame;  Check whether there is undercoverage, ineligible units, duplication	Develop clear specifications of the target population;  Try to locate a readily available list that includes as much of target population as possible;

# To reduce sampling error

Indicators of Quality Data	Quality Control Procedure	Quality Assurance Procedures
1). Reasonable margin of error;  2). Representativeness of sample by known demographic parameters	1). Make sure the margin of error is below 5% for expected number of respondents;  2). Compare the distributions of obtained sample and those of the population by certain demographic characteristics  Check whether key subgroups are represented in the sample	Calculate a reasonable sample size based on the size of the sampling frame, expected margin of error, and anticipated response rate, and the resources available for the survey;  Correct implementation of sampling procedure

## With a sampling frame of 10000 students

Expected # of respondents	Margin of Error (95% confidence level)	Anticipated Response Rate	Sample Size
300	5.57%	20%	1500
400	4.80%	20%	2000
500	4.27%	20%	2500
600	3.88%	20%	3000
700	3.57%	20%	3500
800	3.32%	20%	4000
900	3.12%	20%	4500

# To reduce non-response error

Indicators of Quality Data	Quality Control Procedure
<p>At the unit level:</p> <ol style="list-style-type: none"> <li>1). A reasonably good response rate;</li> <li>2). Insignificant difference between respondents and non-respondents</li> </ol>	<p>Assess unit non-response bias:</p> <ol style="list-style-type: none"> <li>1). <math>\text{Response rate} = I / (I + R + NC + O)</math>            I: a complete interview; R=refusal and break-off, NC=non-contact, O=others (excluding ineligible respondents)            OR <math>\text{Response rate} = I / (S - NI)</math>            I: a complete interview; S: sampled recipients; NI: ineligible respondents</li> <li>2). Assess the difference between the respondents and the nonrespondents by:               <ol style="list-style-type: none"> <li>a). Assess the degree to which non-response will interact with the topics or issues of the survey.</li> <li>b). Compare survey respondents' demographic characteristics to those of the sampling frame</li> <li>c). Examine the characteristics of the late respondents.</li> </ol> </li> </ol>
<p>At the item level:</p> <ol style="list-style-type: none"> <li>1). A reasonable proportion of missing data</li> <li>2). Insignificant difference between respondents and non-respondents</li> </ol>	<p>Do item non-response analysis by</p> <ol style="list-style-type: none"> <li>1). Calculating the proportion of missing data for each question</li> <li>2). Deciding whether the data are missing completely at random</li> <li>3). Investigating the variables with a large proportion of missing data</li> </ol>

# To reduce non-response error (cont'd)

## Quality Assurance Procedures

### **Efforts to reduce non-responses caused by**

#### *Non-contacts:*

- Efforts to reach the recipients

#### *Refusal:*

- Prior notification about the survey;
- Initial contact messages;
- Number and timing of attempts to reach the recipients;
- Data collection period;
- Use of incentives;
- Ways of requesting participation

#### *Inability to participate:*

- Reasonable length of the questionnaire;
- Make sure questions are about relevant, available, accessible or generatable information and avoid asking about inestimable information

# To reduce post-survey error

- Errors in Data Processing and Adjustment

- Correct data processing procedures in:

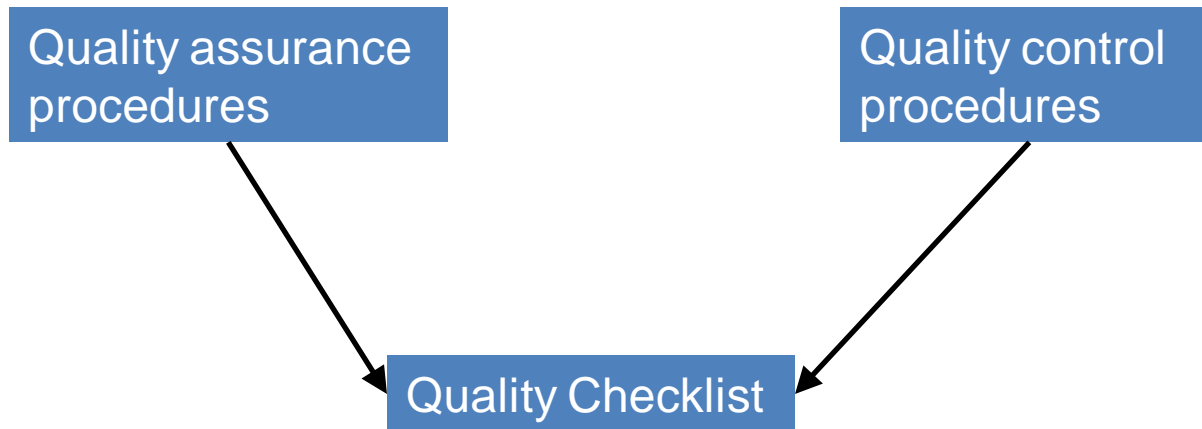
Coding;

Weighting;

Imputation for missing data;

Statistical computing

# Quality Checklist



# DOCUMENTING SURVEY DATA QUALITY

- Set up a data quality file

Definitional

Procedural

Operational

Systems

# An example

## 2009 BCIT Library Services Survey

- Measurement error: validity & reliability
- Sampling error:
  - representativeness of respondents
  - margin of error
- Nonresponse error:
  - response rate;
  - non-response bias

# Concluding remarks

- Solution to improving in-house survey data quality
- Survey data quality: multi-dimensional
- Importance of quality documentation

*Thank you!*

My contact info:

Qin Liu, British Columbia Institute of Technology

qliu@bcit.ca