

How Much Data Do I Need?

Introduction

Participants in this session will be able to:

1. Identify the critical drivers that influence how much data to collect
2. Make choices about data collection based on common tradeoffs
3. Understand the impact data collection choices have on validity and reliability

A Few Data Questions

When conducting an evaluation, the amount of data you need often depends upon the type of data collection needed to answer the evaluation's key questions. Always consider what question you need to answer before choosing a way to answer it.

What do you really want to know?

Who do you want to provide you with it? What happens if you include or exclude people?

When do you need to know it?

What will you do with the information?

How will you maintain integrity and protect your respondents from potential harm?

How Much Data Do I Need?

Definitions

Validity	Measures accurately what it claims to measure
Reliability	Repeatability, consistency in measures across time and context
Population	The overall group to which you'd like to generalize your findings
Sampling Frame	The subset of the population you'd like to target
Sample	The subset of the sampling frame that you want to participate in data collection
Response Rate	The number of people in the sample that actually respond
Confidence Interval	The range of possible results which would contain the result we're looking for (e.g., an employee's satisfaction score will most likely fall between 85-89% for a particular question)
Confidence Level	How likely we are to find the score we're looking for within a particular range (e.g., 95% certain the employee's score will be within 85-89% on that satisfaction question)
Statistical Significance	The results you are seeing are not due to chance, they actually exist because there are true differences
Power Analysis	Determining whether or not you have enough data to be confident that your results are occurring for real, NOT because of chance
Effect Size	The magnitude of difference that exists

How Much Data Do I Need?

Types of Data

Quantitative

Nominal

Ordinal

Interval

Ratio

Qualitative

Type of Data	Issue	Common Tradeoffs
Quantitative	Fast and efficient Less resource intensive Limited descriptive capabilities Anonymity possible	Limited explanatory value Can be misinterpreted Perceived as impersonal Breadth, not depth
Qualitative	Data takes time to collect Rich description <i>How</i> and <i>why</i> become accessible	Resource-intensive High-touch Depth, not breadth Confidentiality risks

Notes:

How Much Data Do I Need?

Data Considerations

	Critical Drivers	Notes
Quantitative	<ul style="list-style-type: none">• Desired statistical significance levels• Confidence levels & intervals• Representativeness of population• Diversity of opinion• Data collection methods	
Qualitative	<ul style="list-style-type: none">• Choice of sources (business & social factors)• Resource availability and funding• Time	
Data Sources	<ul style="list-style-type: none">• Representativeness and diversity• Validity = accuracy, truth• Reliability = repeatability of results• Triangulation• Integrity (especially security and bias)• Characteristics – network hubs, gatekeepers, politics, culture, access• Perishability – time- and condition-sensitive	

Better to have more detail that you can consolidate than less detail that you can't separate.

How Much Data Do I Need?

Case Examples

Customer Satisfaction

Your organization decides it wants to know how satisfied your customers are with the service they are receiving. What kinds of data might help you identify what that satisfaction includes? What datasources would give you an accurate picture of satisfaction?

Employee Engagement

HR wants to conduct an annual engagement survey in addition to measuring employee satisfaction. What would you want to know about how engagement is measured?

Your Organization

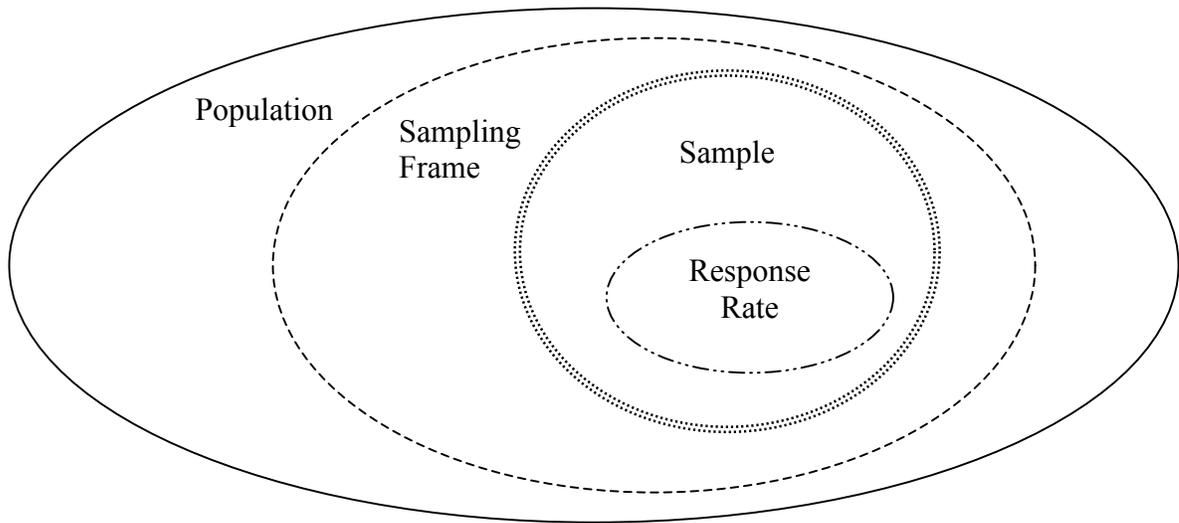
What new initiative of data collection is your organization pursuing? What questions would you be likely to ask now that you have a better understanding of types of data, validity, and reliability?

How Much Data Do I Need?

Sample Size Estimations

- Use a table of participant totals for confidence percentages (widely available on the Internet)
- Sample Size \neq Response Rate

Example: For a 95% confidence level for a survey, if your *population* size is 70 managers, you need a *sample* of at least **59** to complete your survey (that's an 84% response rate if you can get it). If only 30 reply, you cannot be confident in your results with only a 43% response rate.



Statistical Power

You may be able to get enough people to answer a survey to have high confidence in the results. That assumes that your survey has already proven to measure accurately what you want it to measure. If you're not sure, you have to establish validity. What does that take?

Example: You create a competency model for your technical workers. There are 10 competency domains that cover a variety of tasks and topics. Let's say you have between 5 and 8 questions per domain that assess the competency. To do a factor analysis (exploratory factor analysis) to ensure your competency questions accurately represent the domain (establishing validity), you need at least 20 responses as a rule of thumb. How much data is that? 20 responses x 5 questions x 10 domains = 1000 unique bits of data. You need to have a minimum of 1000 respondents. This isn't practical for many organizations unless you commit to a multi-interval data collection process over time or have a lot of employees. This is why you need to think about validity and reliability and sampling before collecting data.

How Much Data Do I Need?

Resource List

- Binder, C. (2010). Measurement, evaluation, and research: Feedback for decision-making. In J. Moseley and J. Dessinger (Eds.), *Handbook of Improving Performance in the Workplace: Volume 3 – Measurement and evaluation*, 3-24.
- Dillman, D., Smyth, J., & Christian, L. (2009). *Internet, mail, and mixed-mode surveys: The tailored design method*. 3rd Ed. Hoboken, NJ: John Wiley & Sons.
- Dobrovolsky, J., & Fuentes, S. (2008). Quantitative versus qualitative evaluation: A tool to decide which to use. *Performance Improvement*, 47 (4), 7-14. DOI: 10.1002/pfi.197
- Gall, M., Gall, J., & Borg, W. (2003). *Educational research: An introduction*. 7th Ed. Boston: Pearson.
- Henry, G. (1990). *Practical sampling*. (Vol. 21). Newbury Park: Sage.
- Kraemer, H.C., & Thieman, S. (1987). *How many subjects?* Newbury Park, CA: Sage.
- Nardi, P. (2006). *Doing survey research: A guide to quantitative methods*. 2nd Ed. Boston: Pearson.
- Russ-Eft, D., & Preskill, H. (2009). *Evaluation in organizations: A systematic approach to enhancing learning, performance and change*. 2nd Ed. New York: Basic Books.

Contact Information:

Stephanie Christine G. Fuentes, Ph.D.
Inventivo Design, LLC
<http://www.inventivodesign.com>
PO Box 281068
Lakewood, CO 80228
1.720.413.0600
stephanie@inventivodesign.com

How Much Data Do I Need?

Presenter Notes:

All data collection is a snapshot in time. Repeated data collection can help identify trends over time. Caution should be used when making generalizations about results given their snapshot nature.

If you ask the wrong question, no data collection method can fix that. Be smart about what you ask and how you ask it.

Put together a critical driver and a tradeoff – for example, you might choose only a few really knowledgeable people to provide rich data, but it's very expensive and may not capture the breadth of the knowledge available. Is that what you want?

Estimating how much data you need for common data collection activities

Sample size versus response rates

Oversampling for Representativeness

Estimating sample size based on significance levels (see [EDPSY notes](#), or [article draft](#))

Common sense approach – would I be okay with this?

Impact of data collection choices on reporting and resources

You can't say it's valid when...

Only 10 percent of the desired population responds.

Your data collection process lacks integrity.

Responses were biased by external issues (incentives & disincentives)

Watch out for influential datasources. (example: You didn't contact ME, therefore it's invalid)

Bias and politics in reporting

Expect bias – find a way to reduce it. (example: provide anonymity)

Expect interference – stand your ground to maintain the integrity of the results.

Cautionary notes – these are guidelines, not rules.

Control your reporting as PDFs and involve clients with caution – you don't want the report to be substantively different, or skewed to serve an agenda

Presenting unwelcome findings – your ethical obligation.