Social Science Survey Data Sets in the Public Domain: Access, Quality, and Importance

David Howell
dahowell@umich.edu

The Philippines
September 2014
Presentation Outline

• Introduction
• How can we evaluate survey quality?
• Dataset and archive examples
• Data use, training, and citation
• Benefits of data sharing and collaborative data collection
First, my thanks!

...a private and independent academic institute which conducts survey research on topics of public interest

http://www.sws.org.ph
Who is this guy, anyways?
• Quantitative research, training, capacity building, and public goods production

• Interdisciplinary
  Political Science – Social Work – Economics
  Natural Resources – Education – Public Policy
  Statistics – Complex Systems – Communication
Associated with:

– American National Election Studies
– Arab Barometer
– Comparative Study of Electoral Systems
– Constituency-Level Elections Archive
– Empirical Implications of Theoretical Models
– Longitudinal Study of American Youth
– World Values Survey
David Howell

• 20+ years at Institute for Social Research
• Currently working on
  — Comparative Study of Electoral Systems
  — International research capacity building
  — Program evaluation
  — Research project development
• Previously
  — American National Election Studies
  — Health and Retirement Study
How can we evaluate survey quality? (some of the issues to consider)
Transparency

Does the study provide documentation of its...
  ...
  ...
  ...
  ...

CSES Module 1 Election Study Archive

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>English</td>
<td>Jul-26-2002</td>
<td>Jul-26-2002</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Untranslated</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Reliability and validity

A measure is **reliable** to the extent that it gives the same result again and again if the measurement is repeated.

A measure is **valid** if it actually measures what it purports to measure.

- W. Phillips (Phil) Shively
In the questionnaire design process:

• Were questions validated on previous surveys?
• Were new questions tested in cognitive interviews? Focus groups?
• Were reliability tests applied to new questions?
• Were questions pretested? On a similar population?
• Do items relate to other measures as expected?
Translation

If translation was used:

- Who translated the questionnaire?
- Was the translation checked or evaluated?
- Was the translated questionnaire pretested?
- What problems were there in doing the translation? Did all concepts translate well?

- adapted from ISSP/Janet Harkness
Sampling Procedures

- Is it a probability sample at all stages, or are there non-probability components?
- Is there detailed documentation of the design and procedures for all stages (selection methods, probabilities, clustering, etc.)? Are quotas or replacement used?
- Are there coverage issues (geography, demographic, language, or technology)?
- Is the sample size adequate to represent the population and sub-populations of interest?
Mode of Interviewing

Is the mode used appropriate for the study?

- Face-to-face
- Telephone
- Mail/Self-Completion
- Internet
- Mixed mode

<table>
<thead>
<tr>
<th></th>
<th>CSES</th>
<th>Module 1</th>
<th>Module 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Face-to-face</td>
<td>70%</td>
<td>71%</td>
<td></td>
</tr>
<tr>
<td>Mail/self</td>
<td>15%</td>
<td>7%</td>
<td></td>
</tr>
<tr>
<td>Telephone</td>
<td>10%</td>
<td>10%</td>
<td></td>
</tr>
<tr>
<td>Mixed</td>
<td>5%</td>
<td>12%</td>
<td></td>
</tr>
</tbody>
</table>
Field Practices

• Have interviewers received General Interviewer Training (GIT), and training in the administration of the specific survey?
• What efforts were made to achieve a high response rate? What was the response rate? Was there differential non-response?
• Was refusal conversion practiced?
Cross-National Issues

Measurement equivalence/comparison:
• What methods were different across countries?
• Were questions comparably implemented?
• Do the questions make sense in all of the participating countries? Culturally appropriate?
• Were any questions sensitive or not legally allowed in certain countries?
• Were there any constraints on respondents to be able to answer freely?
Post-Survey Processing

Were the data processed and documented?

- Examine distributions, relationships of variables
- Skip pattern, missing data, outlier checks
- Weights for sample, non-response
- Are all codes, meanings, etc. identified?

Is there a mechanism for informing users of further errors when identified?
Are problems documented?

Surveys are real world efforts with real world constraints. No survey is perfect.

The imperfections of a study should not be hidden, but highlighted.

— Enhances credibility of project
— Improves the quality of resulting analyses
Show me the data!
(some dataset and archive examples)
Comparative Study of Electoral Systems

- Founded in 1994
- The CSES Module is included in high quality national post-election surveys around the world
- A new substantive theme and questionnaire (with some repeating questions) every five years
- Micro-macro design, to study variations in electoral systems (and other political institutions) and impact on individual attitudes and behaviors
Comparative Study of Electoral Systems

• Module 1 (1996-2001)
  — System performance

• Module 2 (2001-2006)
  — Accountability and representation

• Module 3 (2006-2011)
  — Electoral choices

• Module 4 (2011-2016)
  — Distributional politics and social protection
  — Mobilization
Comparative Study of Electoral Systems

Most common dependent variables across modules:

- Economic voting
- Voter turnout
- Citizen Engagement/ Efficacy
- Satisfaction with Democracy
- Government accountability
- Party Systems/ Cleavages
- Choice parameters
Comparative Study of Electoral Systems

Module 3 Collaborators Map
(not all were successful in collecting data)
Comparative Study of Electoral Systems

• 140 election studies so far, in 50+ countries, including:
• Can download data, or analyze online
• All data are for free and without embargo

Website: http://www.cses.org
Twitter: @csestweets
Global Barometers

• Attitudes and values toward politics, power, reform, democracy and citizens' political actions

• A collaboration among independent studies:
  — Afro Barometer (www.afrobarometer.org)
  — Arab Barometer (www.arabbarometer.org)
  — East Asian Barometer (www.asianbarometer.org)
  — Latino Barometer (www.latinobarometro.org)
...among others

• Online data access:  http://www.jdsurvey.net/gbs/gbs.jsp
East Asian Barometer

• Coordinated out of Taiwan
• Currently available for public access:
  — Cambodia (2008)
  — Indonesia (2006)
  — Singapore (2006)
  — Vietnam (2005)
• Online data analysis: http://www.jdsurvey.net/eab/eab.jsp
Some questions from Philippines 2005:

- Socio-demographics
- Economics of country and family in past, present, future
- Trust in institutions: government, courts, police, media
- Social capital
- Safety and security
- Vote choice, party preference, and political involvement
- Globalization
- Satisfaction with government and democracy
- Most important problems in the country
- International relations
Asia Barometer

- Different from the East Asian Barometer
- Coordinated out of Japan
- Data collection starting in 2003
- East, Southeast, South, and Central Asia
- Website: https://www.asiabarometer.org/
Asia Barometer

Focuses on the daily lives of ordinary people and their relationships to family, neighborhood, workplace, social and political institutions, and market place

- physical, psychological, and sociological dimensions
- affective and cognitive qualities of life
- the types of goods and services they value in order to improve the quality of their own lives and their country's
- prepares for market potentials - developmental, democratic and regionalizing
International Social Survey Programme

• Collecting data since 1985
• 49 participating countries
  — including the Philippines
• Website: http://www.issp.org
• A different topic each year
  — Some repeated in subsequent years
International Social Survey Programme

- Role of government
- Social networks
- Social inequality
- Family, gender roles
- Work orientations
- Religion

- Environment
- National identity
- Social support
- Citizenship
- Leisure and sports
- Health
World Values Survey (WVS)

- Worldwide network of social scientists
- Studying changing beliefs and values and their impact on social and political life
  — economic and technological developments bring major changes in people's values, beliefs and behaviors
World Values Survey (WVS)

Additional topics, among many more:

— Understanding democracy
— Globalization and gender
— Culture, diversity, and religion
— Happiness
— Trust, civic norms, and civil society
— Repression and legitimacy
World Values Survey (WVS)

Countries included in the WVS data archive
World Values Survey (WVS)

- Surveys from 1981 to the present
- Over 450,000 respondents
- Data from 100+ countries, covering nearly 90% of the world’s population
  - Malaysia (2006, 2011)
  - Thailand (2007, 2013)
World Values Survey (WVS)

- Can download data or analyze online
- Website: http://www.worldvaluessurvey.org
- Twitter: @ValuesStudies
Data Archives: International

- ASEP/JDS Data Bank
  http://www.jdsurvey.net/

- ICPSR
  http://www.icpsr.umich.edu/
  Twitter: @ICPSR

- GESIS – Leibniz Institute for the Social Sciences
  http://www.gesis.org/
Data Archives: International

• Roper Center at University of Connecticut
  http://www.ropercenter.uconn.edu/
  Twitter: @RoperCenter
  ... also includes a questionnaire bank

• UK Data Archive
  http://www.data-archive.ac.uk/
  Twitter: @UKDataArchive
National Statistics Office (NSO) Data Archive
http://www.census.gov.ph/nsoda/index.php/home

- Formed in 2009
- Archives large-scale surveys from health and research organizations
- Datasets from 1991 to present
Some of the surveys in the National Statistics Office (NSO) Data Archive:

- Monthly Integrated Survey of Selected Industries
- Quarterly Labor Force Survey
- Annual Family Income and Expenditure Survey
- Annual Family Planning Survey
- Annual Survey on Overseas Filipinos
- Annual Survey of Retail Prices
Data Archives: Philippines - SWS

Social Weather Stations (SWS) Data Bank
http://www.sws.org.ph

- At least partially fee-based
- Hundreds of national and sub-national surveys of the Philippines, 1984 to present
- Many additional foreign surveys as well
Data use, training, and citation
Data use: offline

• Download to your local computer
• Analyze in:
  — SAS, SPSS, STATA (for fee software)
  — R (free software)
• Good for:
  — in-depth analysis
  — creating derivative files
  — replication
Data use: online

- Does not require downloading or local software
- Limited analytical capability
- Good for:
  - exploration
  - non-analysts
  - those without software
Data use: online – WVS example

### V11.- State of health (subjective)

#### Cross by

<table>
<thead>
<tr>
<th></th>
<th>Number of cases</th>
<th>%/Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very good</td>
<td>213</td>
<td>17.7%</td>
</tr>
<tr>
<td>Good</td>
<td>466</td>
<td>38.8%</td>
</tr>
<tr>
<td>Fair</td>
<td>460</td>
<td>38.3%</td>
</tr>
<tr>
<td>Poor</td>
<td>62</td>
<td>5.1%</td>
</tr>
<tr>
<td>(N)</td>
<td>(1,200)</td>
<td>100%</td>
</tr>
</tbody>
</table>

Selected sample: Philippines 2012 (1200)

#### Bar Chart

State of health (subjective)

- Very good
- Good
- Fair
- Poor

Percent of Base N= 1200

#### Question wording

All in all, how would you describe your state of health these days?. Would you say it is:

#### Equivalences in other waves

1981/1984: v11
Data training

• Many useful resources and tutorials can be found on the Internet (for free)
• Learning through collaborating
• University courses are a good option
• There are also many summer institutes
  — A good opportunity to learn methods while networking with global peers
  — Many offer a limited number of scholarships, including to international participants
Some examples of summer institutes are...
Data training: summer institutes

Summer Institute in Survey Research Techniques

• at University of Michigan
• June-July
• focus on surveys: design, data collection, and analysis

http://si.isr.umich.edu/
Data training: summer institutes

Summer Program in Quantitative Methods of Social Research

- at University of Michigan
- mid-June to mid-August
- focus on data analysis (not just surveys), basic through advanced

http://www.icpsr.umich.edu/icpsrweb/sumprog/
Data training: summer institutes

Summer School in Social Science Data Analysis

- at University of Essex
- July-mid August
- focus on data analysis (not just surveys)

http://www.essex.ac.uk/summerschool/
IPSA-NUS Summer School for Social Science Research Methods

- At National University of Singapore
- June
- Quantitative, qualitative, and formal methods

http://methods-school.nus.edu.sg/
Data citation

• Please cite data! Just as you would a paper.

• This helps research projects:
  — become better known
  — understand how their data is being used
  — improve their product
  — raise funding

• Most public data sets have citation guidance:
  — On their website
  — In their documentation
Benefits of data sharing and collaborative data collection
Who benefits from data sharing?

— Other scholars/researchers
— Students
— Policy makers
— Governments
— Non-governmental organizations
— Journalists
— Data collection organizations
— The general public
Data Sharing Benefits

• More visibility for your effort
• Your data get more use, and in more ways
• Can be used in teaching students
• Data have more impact, making funders happy
• Scientific integrity: replication of results
• Preservation and archiving
• Credibility for your organization
• Funding agencies increasingly encouraging and/or requiring the sharing of data
  — Differs by country
  — Differs by discipline
  — Differs by funding source
• Examples in the United States
  — National Institutes of Health
  — National Science Foundation
Collaborative Data Collection Benefits

• Networking with colleagues of similar interest
• Learning new skills and methods
• Presentation and publication opportunities
Collaborative Data Collection Benefits

- Participation in governance of the project
- Promotion of your organizational capacity
- Advice and collaboration in raising funding

Cross-national participation benefits:
- International profile for your country
- Comparative analysis
Questions?

Thanks for your time!

David Howell
dahowell@umich.edu