Correlates of the Item Non-Response in Survey Research: Analysis of the KGSS Cumulative Data, 2003-7

Sang-Wook Kim
(Survey Research Center, Sungkyunkwan Univ., Korea)
The Problem

- Major criteria for the ‘quality data’
  - ✔ Unit non-responses (valid response rates)
  - ✔ Item non-responses (response completeness) \( \leftrightarrow \) DK’s
  - ✔ Psychometric properties (measurement reliability and validity)
  - ✔ Atypical response patterns (ex. response sets)

- Purpose of the Study: Uncover the correlates of DK’s by figuring out how DK’s vary by some of the most important characteristics of the respondents and survey setting
Anatomy of Item Non-Responses (DK’s)

- **Typology & Definition**
  - ✔ Unit non-response: Total failure to obtain any measurement from a given case in the sample due to non-contact, refusal, and inability to participate
  - ✔ Item non-response (DK): Partial failure to obtain valid or usable observation values for some part of a successfully completed case

- Variety of ‘Missing Data’
  1. System-missing (NA): Inapplicability or ineligibility of a certain item (ex. abortion experience) to some group of respondents (ex. males)
  2. Non-system-missing
     1. Process item non-response: Missing data resulting from problems w/ pre-survey questionnaire formatting (ex. improper skip pattern) or post-processing procedures (ex. editing, coding, and punching) → Detected and minimized by routine computerized checks
     2. Interview item non-response: Missing data resulting from social, psychological, and cognitive dynamics of the interview → Common meaning of ‘missing data’

☞ Focus of this study is on non-system missing, interview item non-responses
Impacts of DK on Substantive Research

Exclusion of DK’s from the analysis w/out a closer inspection or serious consideration leads to two sorts of problems
(1) Increased sampling error (due to the reduced N)
(2) Bias in statistical estimation (due to the enlarged deviation of the expected value of sample statistics from population parameters)

- Two conventional remedies of DK’s
  (1) Data imputation
  (2) Survey design
  ➤ Survey design is more important!
Meaning or Interpretation of *DK*

- Are *DK*’s *true underlying values*?
  - Yes! (Ex.) Choice of *DK* for voting preference implies that *R* is undecided yet
  - If Yes, *DK*’s do not automatically qualify as item non-responses
  - No! Because it’s almost impossible in practice to determine if *DK*’s represent true underlying traits
  - If No, *DK*’s qualify as item non-responses

- *DK* = ‘Catch-all’ concept that includes so many off-scale responses, such as *can’t choose, not sure, not enough information to form an opinion, no opinion, question mark*, etc. (Francis and Busch, 1975)

- Reasons for *DK*’s
  - Non-attitude (*really don’t know*)
  - Ambivalent attitude (*neither for nor against the issue*)
  - Editing socially desirable answers (*don’t want socially undesirable opinion*)
  - Sacrificing (*don’t have time or don’t want to think about*)
  - Question ambiguity (*don’t know what it means*)
  - Unavailability of mid-point (*no neutral point*)
Correlates of *DK*

*DK’s* are non-random phenomena (Smith, 1984; Krosnick and Fabrigar, 1997)

- Two Overarching Groups of Correlates
  1. **Individual-level correlates**
     - Either cognitive correlates (R’s cognitive ability)
     - Or social correlates (R’s normative orientation)
  2. **DK** is a function of cognitive comprehension and normative motivation (Young, 2001)

- Gender; Age; Educational Attainment; Employment Status; Occupation; Household Income; Residential Area

- Those who provide more *DK’s*, on average, are females, older, less educated, unemployed, who have lower-skill occupations, who have lower income, and rural residents
  - Because they tend to lack cognitive comprehension and normative motivation
(2) **Survey-level correlates**

→ Of particular interest to this study is interview characteristics in the process of fielding.

✔ No of days elapsed to complete the interview; No of visits required to complete the interview:
  Longer days = R finally conceded to the interview request after having been reluctant to cooperate for a while

☞ Two interpretations: (a) R’s lowered motivation → More *DK’s*
  (b) Inaccessibility → Less *DK’s*, since R’s who possess the two meta-components are less accessible

✔ Technique or mode of questionnaire administration: Interviewer-administration generates fewer *DK’s* than self-administration (Groves et al., 2002)

✔ Provision of incentives to the interviewees: Provision of incentives (gift voucher or in-kind goods) will result in lower *DK’s* than non-provision
Methods

- Data: Cumulative five-year rounds (2003-7) of KGSS
- Sampling: Multi-stage area probability sampling
- Structured face-to-face interviews

Selection of cases
- Each year’s data = Main file (replicating core, ISSP module, EASS or special module) + Supplementary file (visit records, interviewing records, administration technique, incentive provision, interviewer characteristics, R’s attitude, etc.)
  → The two files were added together by matching, and the resultant files for five years were then merged across the years
- 1,315 + 1,312 + 1,613 + 1,605 + 1,431 = 7,276 – 67 (incomplete info for supplementary data ) = 7,209 (N) → Listwise deleted N = 6,839

Selection of items
- Only items that are commonly included throughout the five years
  → Makes it possible to keep maintaining exactly same no of cases across the items
- Only items that have identical wordings throughout the years
- Only items that are not filtered at all by any previous items
  → Inclusion of filtered items makes the no of cases across all items fluctuate
  → A total of 48 items
• Measurement: (Table 1)
  ✔ *DK Rates* = Summated measure created by counting, for each case, the occurrence of *DK* responses across a list of 48 items
  ✔ Three different summated measures of *DK* rates
    (1) *DK* rates for total items (48)
    (2) *DK* rates for behavioral items (16)
    (3) *DK* rates for attitudinal items (32)
    ➔ To cope w/ the skewed dist, all the three measures were log-transformed

• Analysis
  ✔ Correlation and OLS regression analysis
    (Note) 1. Multicollinearity assumption checked (see Table 2, too)
    2. Will refrain, as much as possible, from putting much emphasis on the statistical significance due mostly to a large *N*
Results

- **Univariate Analysis** (Table 1)
  - ✓ *DK* rates for behavioral items (.1993) are far lower than those for attitudinal items (1.1412)
  - ✓ Some correlates, such as years of schooling, occupation, and household income, turn out to contain more *DK*’s than others

- **Bivariate Analysis** (Table 2)
  - ✓ *DK* rates for behavioral items are only weakly correlated w/ those for attitudinal items
  - ✓ Directions each individual- and survey-level correlate is related to *DK* tend to be consistent w/ predictions in this study
  - ✓ Magnitude of correlations tends to be substantially larger for attitudinal than behavioral items
Multivariate Analysis (Table 3)

- Overall, the results tend to be similar between total items and attitudinal items, but they tend to be somewhat different between behavioral and attitudinal items.
- The most salient correlates include educational attainment, household income, and administration technique, especially for total and attitudinal items.
- In general, individual-level correlates turn out to be better predictors of DK rates than survey-level correlates.
- The correlates specified in the equation account for some moderate amount of variance in total (23.9%) and attitudinal (21.8%) items, but much smaller amount in behavioral (7.4%) items.
- Some correlates (i.e., employment status, residential area, and no of days taken to complete the interview) turn out to have conflicting, sign-flipping, impacts on DK rates for behavioral and attitudinal items, respectively.
DK responses are indeed a function of the two over-arching, respondent-level, meta-components of cognitive comprehension and normative obligation.

Higher DK rates are consistently observed among those who are less sophisticated cognitively, on the one hand, and who fail to feel socially obliged to offer substantive answers, on the other—i.e., females, older, less educated, unemployed, lower incomer, and rural residents—often irrespective to the type of question items and also often irrespective to the introduction of controls for other correlates.

This finding is consistent with a number of studies in the U.S. and suggests that the demographic groups of people who give off DK in Korea tend to be basically similar to those in the U.S., thereby supporting for the cross-cultural generality of extant studies.
(II) Not merely are $DK$’s a function of respondent-level characteristics, they are also a function of survey characteristics

✔ Higher $DK$ rates are observed when longer days are elapsed to complete the interview, when fewer visits are made to complete the interview, and when self-administration is done

☞ Longer days (and the resultant more $DK$’s) is more likely to be a reflection of R’s lack of motivation, whereas multiple visits (and the resultant less $DK$’s) is more likely to be a reflection of accessibility

☞ Questionnaires lengthy and abundant w/ a lot of complicated items, when allowed to be filled out by R, suffer from the response completeness due either to lack of opportunity to clarify and probe the questions or to R’s lack of motivation to offer all substantive answers
(III) Respondent characteristics, on average, are more salient correlates than survey characteristics, irrespective to the type of question items (behavioral vs. attitudinal)

✔ Particularly salient are the three respondent characteristics of educational attainment, household income, and gender, plus the survey characteristic of administration technique

☞ Importance of the three R characteristics, as proxies for R’s cognitive ability and normative propensity, is well documented in past researches in the U.S.

☞ Further attempts are required to overcome the potential misspecification problem by identifying better or stronger correlates of survey characteristics
The two types of items, behavioral and attitudinal, do indeed behave differently as they relate to *DK* responses and suggested correlates

✔ On average, R’s feel free to disclose some of the factual or behavioral things about themselves (e.g., socio-demographic info, voting behavior, etc.), while they feel more burdensome in expressing their own opinions on some issues (e.g., socio-economic or political orientation, identification, etc.)

✔ The extent to which they feel so, however, varies by some of the individual characteristics: the unemployed and rural residents experience more difficulties in expressing their own opinions on social issues, whereas the employed and urban residents are more reluctant to reveal factual info about themselves

☞ Due either to the meta-components (cognitive comprehension and normative propensity) or the inherently confidential orientation of the cognitively sophisticated and normatively motivated people

✔ The extent also varies by some survey characteristic: the more days are taken to complete the interview, the more *DK*’s there exist for attitudinal items, but the less *DK*’s there are for behavioral items

☞ Motivation factor, rather than accessibility factor, might have been working, which is typical to those who are less sophisticated cognitively and less obliged socially

☞ Support for an attempt to demarcate the two types of items!
Methodological Implications

✔ [Academic Implication] Add something to the understanding of the mechanism underlying DK responses; People are indeed offering DK responses for some identifiable reasons, notably cognitive sophistication and normative propensity

✔ Admittedly, however, we are still left w/ the pending causal mechanism underlying the detailed social and cognitive dynamics leading to DK’s

✔ [Practical Implication] Helps to develop some intervention strategies to minimize DK’s

(1) Survey researchers have to normatively motivate R’s by stressing the importance of the survey or by developing well devised standardized interview protocols

(2) Some prudent surveying strategies targeted mainly for R’s who lack the two meta-components (i.e., female, elderly, less educated, unemployed, lower incomer, and rural residents) could be developed: Ex. More amount of interviewer attention, interviewing time, clarification, probing, etc.

(3) What is urgent for R’s who possess the two meta-components (i.e., male, younger, educated, employed, middle- to high-incomer, and urban residents) is to develop strategies to make them more easily accessible and actually concede to the survey request: Ex. Repeated call-backs in different times and places
Virtue of the Study?

☞ The first full-fledged attempt to figure out the correlates of DK’s in Korea

✔ No argument is made, though, that this study is fully generalizable beyond the KGSS, a survey framework that contains several features unique to it

→ Further studies are thus encouraged in the future
More Information

Available at swkim@skku.edu
or http://www.src.re.kr