Westat

Online probing for questionnaire evaluation: Effects of sample source and analysis method

Reanne Townsend, Rosalynn Yang, Kristin Chen, Gonzalo Rivero, & Terisa Davis (Westat)

Gordon Willis, Stephanie Fowler, & Richard Moser (NIH)

AAPOR 2018 Taking Survey and Public Opinion Research to New Heights

Background and Introduction

- Online Probing (OP) is a questionnaire evaluation methodology which administers probe questions within a web survey to assess targeted items. (see Edgar, Murphy & Keating 2016; Meitinger & Behr 2016).

- Some have experimented with Online Probing procedures to determine whether features such as text box size and probe placement affect data quality (e.g. Behr, Bandilla, Kaczmirek & Braun 2014; Fowler et al 2017).

- However many questions remain about how other features of Online Probing study design may influence results.

Research Questions

1. How does the ***amount and quality of data*** provided in response to Online Probing differ by ***sample source or recruitment strategy***?
   - Probability, nonprobability with quotas, convenience sample

2. How do ***content analysis results*** differ by ***analysis method***?
   - Traditional “hand-coding” vs. unsupervised keyword extraction

Design

- Short 10-minute web survey completed by 3,089 respondents
  - Questionnaire composed of items from the Health Information National Trends Survey (HINTS)
- Respondents come from 3 different web panels, using varied sampling or recruitment methodologies

probability	***nonprobability***	
	quota	***convenience***
GfK (n=1,033) Probability-based sample (ABS/RDD)	**YouGov** (n=1,000) Nonprobability sample with demographic quotas	**mTurk** (n=1,056) Nonprobability convenience sample

Design

- Online Probes were administered as open-ended questions at the end of the questionnaire (retrospectively)
  - One probe each for 4 items. Mix of question and probe types
    - *2 ask respondents to list examples of a construct ("social media", and "medical information")*
    - *1 asks for method for calculating whether smoked 100 cigarettes*
    - **1 asks for reason behind evaluation of cancer likelihood**

Design

Cognitive probe used for thematic analysis

Earlier in the survey, you were asked the following question:

How likely are you to get cancer in your lifetime?

- ☐ Very unlikely
- ☐ Unlikely
- ☐ Neither unlikely nor likely
- ☐ Likely
- ☐ Very likely

You said you were [FILL RESPONSE] to get cancer in your lifetime. Why do you think that?

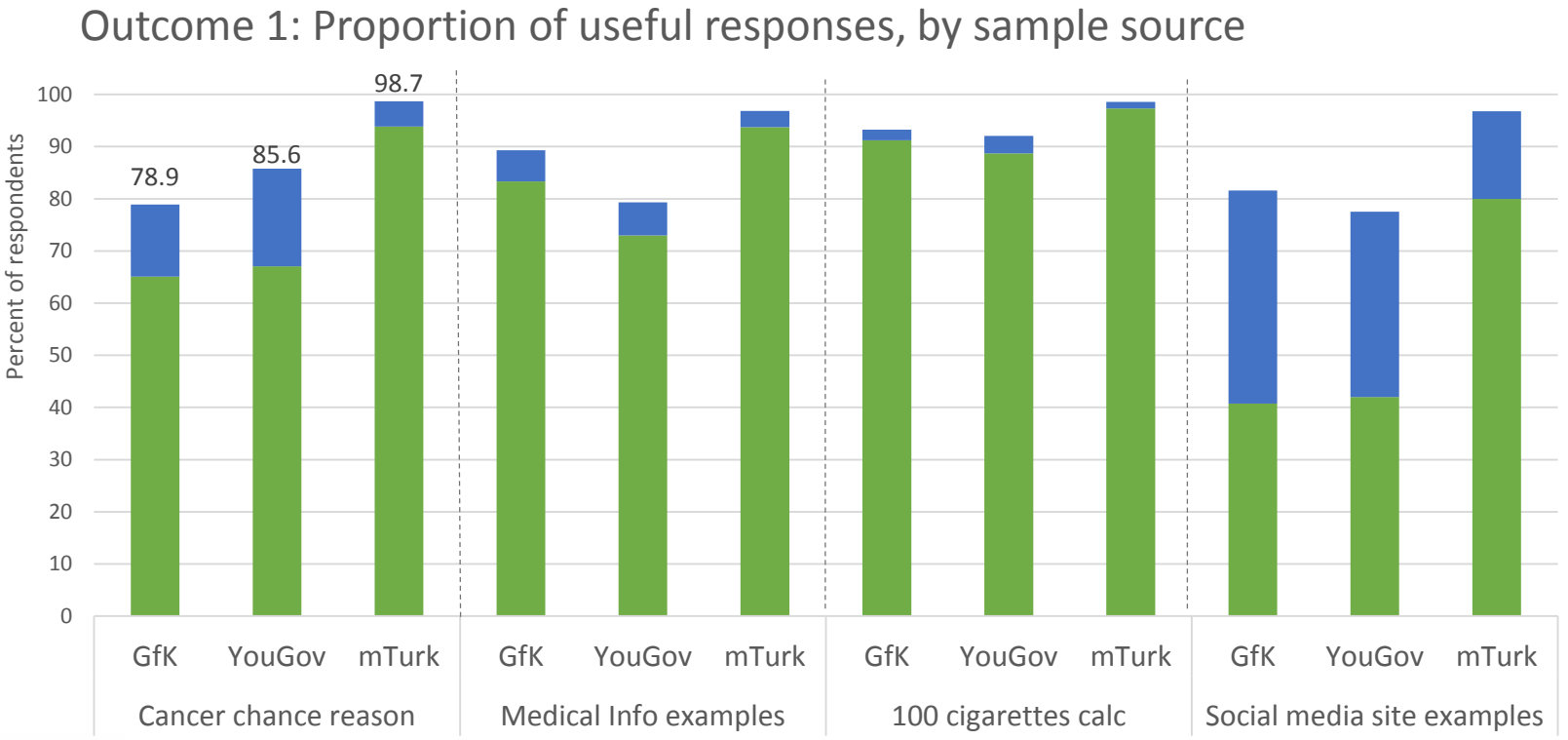
Research Question 1: Sample Source

RQ1 Analysis

How does the amount and quality of data provided in response to Online Probing differ by sample source or recruitment strategy?

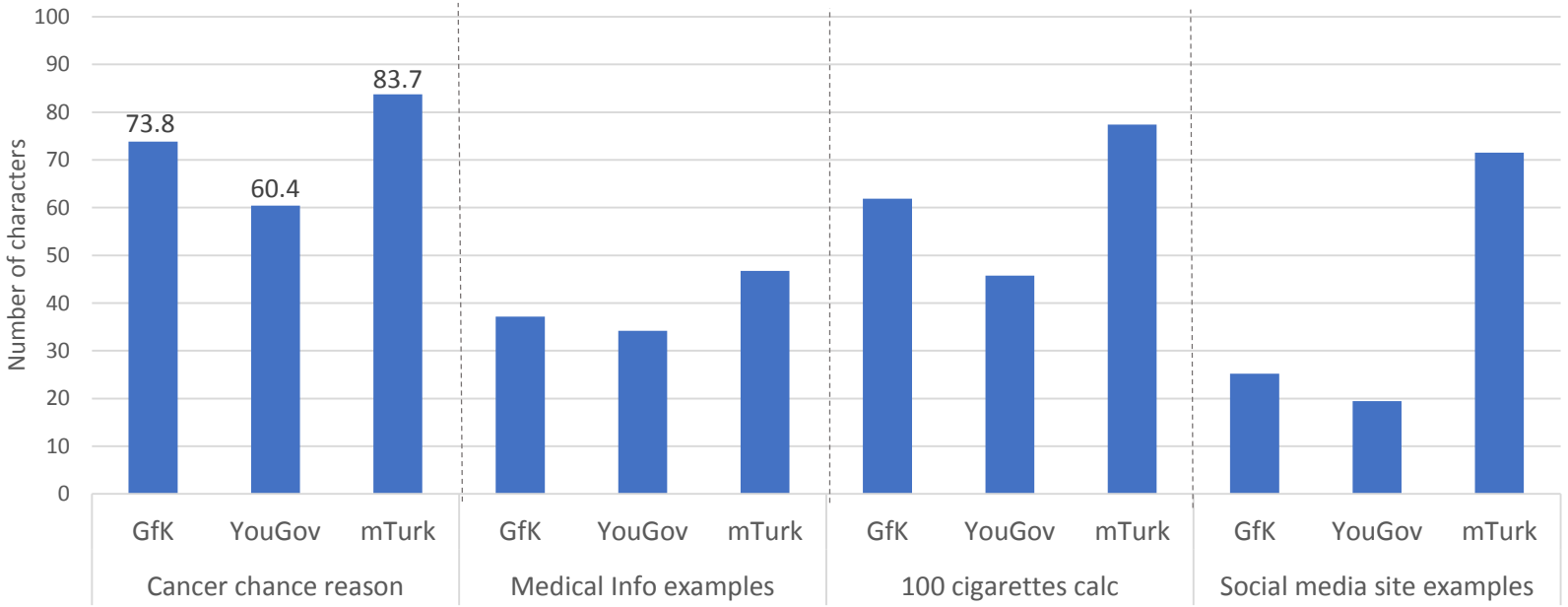
- Outcome 1: Proportion of respondents giving a ***"useful" response***
  - Coded by hand
  - Nonresponse/off-topic, Minimal response, Potentially useful response
- Outcome 2: Average ***character count*** among potentially useful responses
  - Excluding spaces

RQ1 Results


Outcome 1: Proportion of useful responses, by sample source
Percent of respondents
100
90
80
70
60
50
40
30
20
10
0
78.9
85.6
98.7
GfK
YouGov
mTurk
Cancer chance reason
Medical Info examples
100 cigarettes calc
Social media site examples
Useful response
Minimal response

RQ1 Results

Outcome 2: Average number of characters per useful response, by sample source



RQ1 Results

Summary

- mTurk respondents consistently provide longer and more useful responses compared to the other web panels
  - Could be that mTurk workers satisfice less due to the option mTurk requestors have to reject unsatisfactory work (resulting in no payment)
- There is also variance in length and usefulness of responses by type of probe
  - Probes asking for respondent to list examples seem have shorter and less useful responses

Research Question 2: Analysis Method

RQ2 Analysis

How does analysis method affect a thematic content analysis of the open ended web probe responses?

- Method 1: Traditional “by-hand” coding
  - 2 coders, categories determined jointly by coders, all responses coded (75 double)
- Method 2: Natural Language Processing (NLP) – unsupervised keyword extraction and topic model
  - Identify relevant keywords
  - Group keywords into “topics” based on contextual similarity
    - Pre-trained word embedding model Word2Vec, trained on Google News
  - Associate individual responses with topics based on the occurrence of keywords

**note: both methods allowed for multiple categories per response*

RQ2 Results

Results of Thematic Coding using traditional, by-hand method

Category	Description	% of responses	Inter-rater agreement
Family	Family History, genetics	**47.2**	1.00
Lifestyle	Smoking, Lifestyle & environment (incl. diet, exercise, pollution, "chemicals")	**31.0**	0.93
Random	Can't know, no way to know, 50/50 chance, can't control it, it's random, it's luck of the draw	**19.7**	0.62
Common	Cancer is common, everyone gets it, everything causes cancer	**12.5**	0.79
Don't know	Don't know/No idea, Don't care, not concerned, don't think about it, why worry	**7.4**	0.60
Other	All other responses (e.g. current age, other health issues, medical advances)	**7.6**	0.58
Faith	Faith, feeling, intuition, positive thinking	**4.6**	0.65

RQ2 Results

Results of Thematic Coding using unsupervised NLP model

Category	Example keywords	%
Family	- parent, ancestor, grandparent, family, sibling, uncle - baby, man, woman, teenager, friend	**37.9**
Belief/Certainty	- luck, presume, uncertain, unsure, hunch, gut, prediction, hopeful, hope - paranoid, everyone, anybody, anytime, jesus, christ, optimism, god, faith	**34.2**
Sun & Other	- sunshine, sun, sunny, beach - environment, industry, metal, research, knowledge, capability, technology, future	**30.3**
Disease, age, lifestyle	-disorder, death, disease, insurance, sick, treatment, condition, lifestyle, longevity -prostate, stomach, heart, freckle, skin, colon, lung, bone, testicular, depression	**15.4**
Actions	take, try, address, counteract, visit, focus, avoid, help, prevent, maintain, protect, exercise, combat, minimize, limit	**13.6**
Risk & fear	-prone, cause, trigger, culprit, precursor, tendency, predisposition -fear, risk, danger, paranoia, harmful, damage, chemtrails	**10.5**
Diet & Smoking	- eating, sugar, vegetable, nutrient, pollution, additive, toxic, chemical - smoker, cigarette, drinker, substance	**6.8**

RQ2 Results

Compare conclusions between analysis methods

- **Similarities**
  - Family history and genetics as most common response
    - 47% of hand coded responses, 38% of NLP responses
  - Many respondents feel they can't predict or control whether they get cancer
    - "Random" and "Faith" from hand coding (24%), "Belief/Certainty" for NLP (34%)

RQ2 Results

Compare conclusions between analysis methods

- **Differences**
  - "Environmental & Lifestyle factors"
    - Hand-coding grouped all lifestyle factors together (incl. smoking, diet, exercise, pollution, sun)
    - NLP has "action", "diet/smoking", "sun & others", and "Disease, age & lifestyle"
      - A lot of overlap with "Environmental & Lifestyle", but not completely
  - "Cancer is common" sentiment did not show up as a category in NLP analysis (13% in hand coding)

Discussion

RQ1. Amount and quality of data by sample source

- The amount and quality of information elicited from Online Probing can differ depending on the source of the sample
  - mTurkers provide more information, but are they “professional respondents” and not generalizable?
- Possible next steps:
  - Examine whether thematic coding results differ by sample source
  - Further exploration of how question and probe type affect amount and quality of information

Discussion

RQ2. Thematic coding results by analysis method

- Thematic categories defined by keywords that can be identified outside of syntactical context can be similar between hand coding and NLP (e.g. family & genetics)
  - Concepts which require context outside of individual keywords are not as easily categorized by unsupervised keyword extraction (e.g. “Sun & Others”)
  - Possible next step: Classification could be improved by using more sophisticated NLP methods, such as using n-grams instead of single keywords, and a probabilistic framework rather than deterministic

Thank you!

Contact:
ReanneTownsend@Westat.com