GfK Knowledge NETWORKS

Technique to Blend Probability and Non-probability Internet Samples

Webinar Presentation – March 1, 2012

Webinar presenters

Charles DiSogra
Chief Statistician
GfK - Knowledge Networks, Palo Alto, CA

Curtiss L. Cobb
Director, Survey Methods Group
GfK - Knowledge Networks, Palo Alto, CA

Session outline

I. Purpose of session
II. Probability and non-probability online panel samples
III. Situations for blending panel samples
IV. Early adopter attitudes
V. Technique for calibrating
VI. Quantitative evaluation of bias (4 examples)
VII. Conclusions & Future Research
VIII. Q & A

Purpose of Session

- Demonstrate a calibration technique for blending probability and non-probability Internet samples by leveraging "early adopter" attitudes in the weighting procedure

- Show examples of bias reduction using this technique

- Review future research directions

DiSogra, C. et al. *Calibrating Non-Probability Internet Samples with Probability Samples Using Early Adopter Characteristics*. In **2011 JSM Proceedings**, Survey Methods Section. Alexandria, VA: American Statistical Association.

Types of online Web panels

1. Probability-based panels
2. Non-probability volunteer "opt-in" panels

1. Probability-based Web panels

- Recruited with probability samples (no non-sampled volunteers)

  Area-based, in-person methods

  Random-digit dial (RDD)

  Dual frame samples of RDD with a cell phone component

  Address-based sampling (ABS)

- Panel members have known selection probability
  - Accounted in panel member's base weight
  - All sampling frame units have a non-zero chance of being recruited

- Used by government, academic and non-profit researchers and private companies where generalizable rigor is desired

Probability-based Web panels

- Samples drawn from panel have high completion rates (65-70%)
- Results are accepted as generalizable
- Can calculate prevalence estimates with confidence intervals
- *American Association of Public Opinion Research* recognizes probability-based samples, *ergo* panels recruited as such, as a valid and reliable survey method
- Due to recruitment costs, current panels tend to be of modest size
  - Range: 2,000 to 60,000 members

2. Non-probability opt-in Web panels

- Large, volunteer membership
  - Panel size can be a million or more
  - Fundamentally, these are convenience samples
- Membership consists of people on the Web who joined through
  - advertisements
  - pop-up invitations
  - e-mail marketing
  - aggregator sites
    (e.g., *surveyclub.com, paidsurveyworld.net, getpaidsurveys.com*)
  - member referrals
- Used extensively by market researchers

  Low cost

  Can target defined audiences with member profile data

  Great for finding very rare populations due to very large membership

Non-probability opt-in Web panels

- Recruitment, sampling, weighting methods are not transparent
  - Use of various forms of quota sampling for panel studies
- Survey completion rates are generally low (2-9%)
- Not considered generalizable for prevalence estimates
- Industry organizations, *e.g., Advertising Research Foundation,* set voluntary standards for membership management, quality
  - E.g., minimize professional respondents, multiple panel membership
- Cost-effective for some types of research and researcher tolerance

Advantages of probability Web samples

Web samples have lower cost per completion than RDD or area-based in-person

More rapid results turnaround than RDD or in-person

- Large samples can be reached quickly
- Faster data collection

Web probability samples are more accurate than RDD*

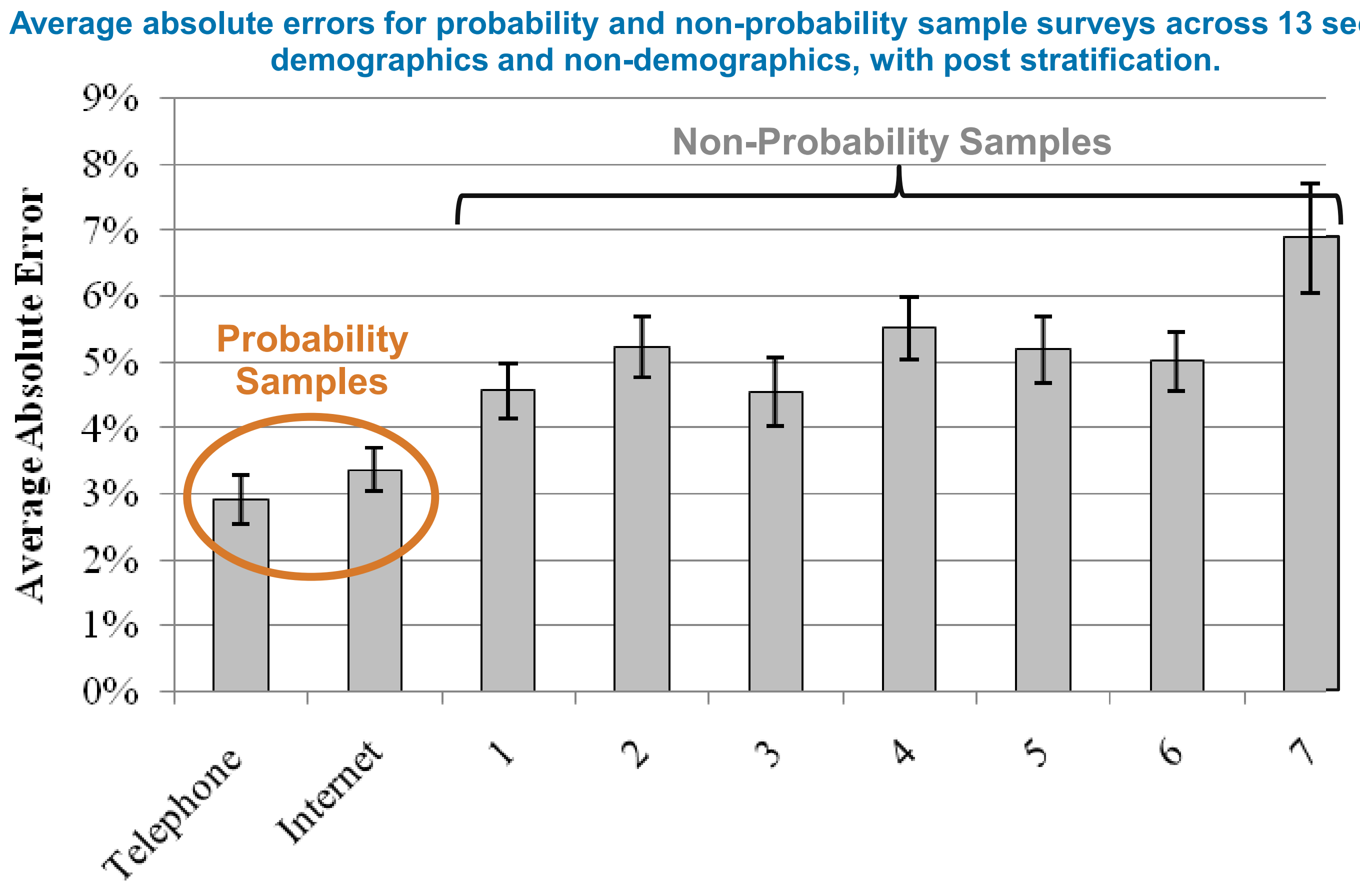
- Higher concurrent validity
- Less survey satisficing
- Less social desirability bias

Probability-based overcomes limitations of opt-in Internet panels

- A known sample frame
- Higher completion rates
- Can reliably generalize findings

* Chang, L., Krosnick, J.A. (2009). National Surveys via RDD Telephone Interviewing vs. the Internet: Comparing Sample Representativeness and Response Quality. *Public Opinion Quarterly 73*: 641-678.

Accuracy of probability Internet and RDD samples

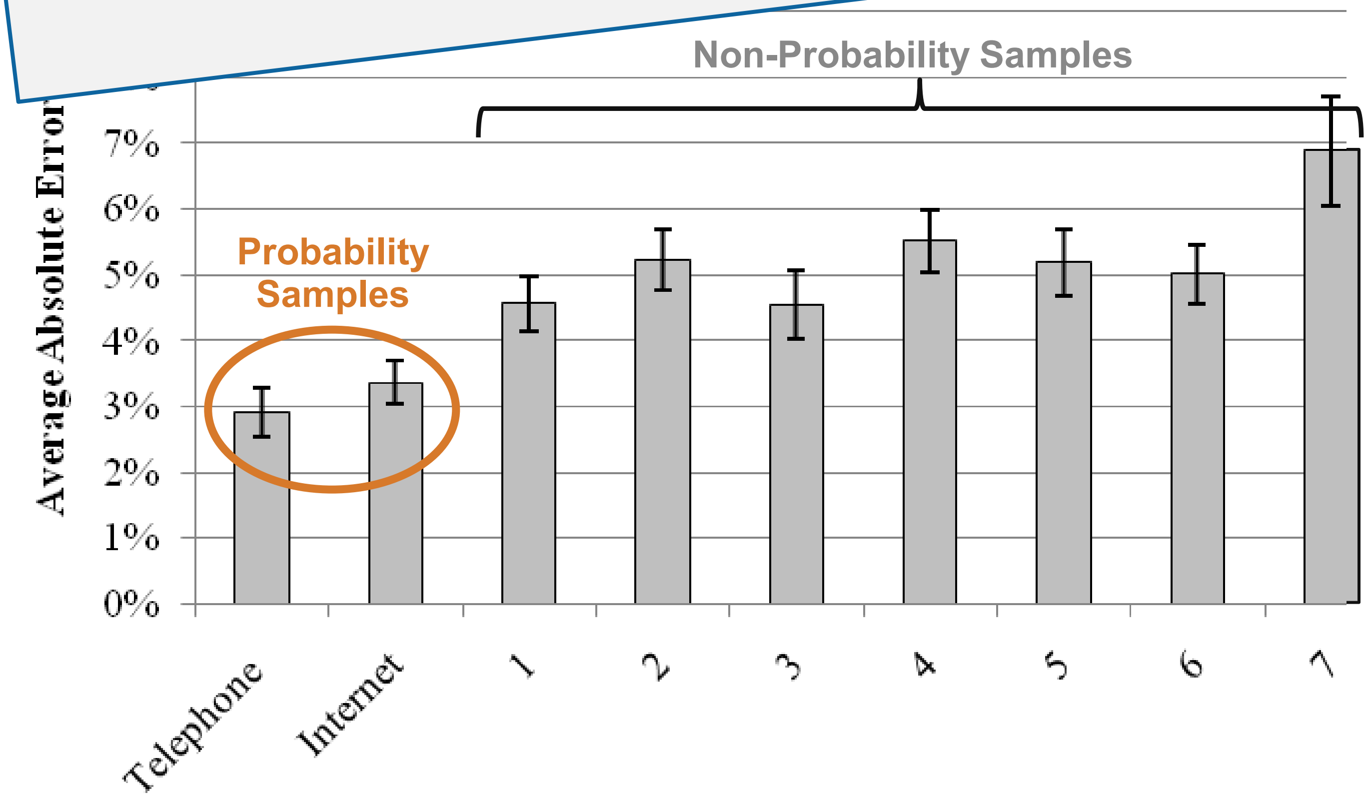


Source: Yeager, Krosnick, et. al., 2011. Comparing the Accuracy of RDD Telephone Surveys and Internet Surveys Conducted with Probability and Non-Probability Samples. Public Opinion Quarterly. 75:709-747..

Accuracy of probability Internet and RDD samples

"Both were significantly more accurate than any of the non-probability opt-in sample Internet surveys."

...ross 13 secondary ... stratification.



Source: Yeager, Krosnick, et. al., 2011. Comparing the Accuracy of RDD Telephone Surveys and Internet Surveys Conducted with Probability and Non-Probability Samples. Public Opinion Quarterly. 75:709-747..

>55,000 members

Probability-based ABS recruitment

Representative of U.S. adult population

Includes:

- Households with no Internet access at time of recruitment
  - 33% of US adults have no Internet access; GfK/KN provides laptop, free ISP
- Cell phone only households (~30% of HHs)
- Spanish-language households
- Extensive profile data maintained on member demographics, attitudes, opinions, behaviors, health conditions, media usage, etc.

12

Situations for blending panel samples

Finite size of probability-based panel

Small or rare populations

- Some examples:
  - Boat owners
  - Recent college graduates
  - Less-common medical conditions
  - Viewers of specific niche cable networks
  - Specific race/ethnic groups when large samples are required

Small geographic area samples

- Some examples:
  - Specific congressional districts
  - Small media markets
  - ZIP code clusters

2-step solution

1. Supplement probability sample with opt-in panel cases

Use quota sampling with opt-in cases to minimize demo skews/weights.

2. Calibrate the combined samples to the probability sample

Use “ancillary information” to minimize bias from the opt-in sample.

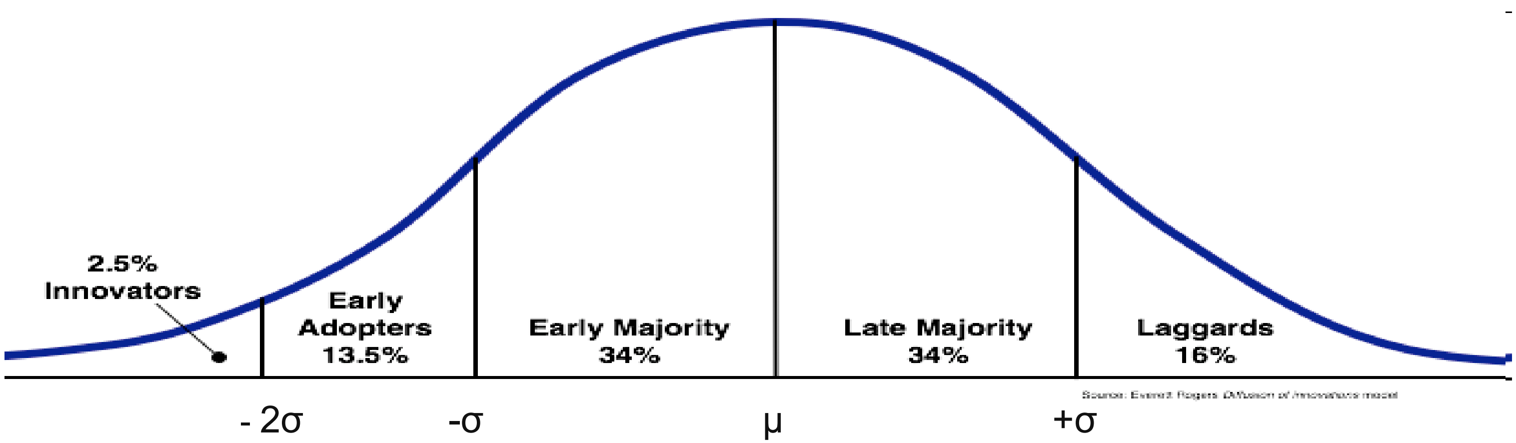
What ancillary information?

Early Adopter Attitudes

Early adopter (*EA*) identity

Consumer research often attempts to identify *EA*s

- EAs embrace new technology/products before most others
- Research goes back to the 1950s
  Bourne, F.S. (1959). The Adoption Process. Reprinted in M.J. Baker's(ed) (2001) *Marketing: Critical Perspectives on Business and Management.* New York: Routledge.



Based on Rogers, M. (2003). Diffusion of Innovations. 5th Edition, New York: Free Press.

Rogers' Normalized Adopter Categories

Early adopter (*EA*) identity

2008 KN Study:

Comparison of *EA* attitudes among Internet panels

(Dennis, Osborn, & Semans 2009)

- Two probability-based panels
  - American National Election Studies (ANES) Web Panel 2007-2009
  - KnowledgePanel®
- Two well-known non-probability opt-in Web panels
  - Web Panel A
  - Web Panel B

Administered the same questionnaire September-October 2008.

Early adopter (*EA*) questions

Percent "Strongly agree / Agree"

Base (n)	ANES Web Panel (1,397)	Knowledge Panel (1,210)	Opt-In Web Panel A (1,221)	Opt-In Web Panel B (1,223)
I usually try new products before other people do.	26.4	24.0	44.2* ↑	41.4* ↑
I often try new brands because I like variety and get bored with the same old thing.	36.6	34.1	52.0* ↑	54.2* ↑
When I shop I look for what is new.	44.5	35.7*	55.2* ↑	59.0* ↑
I like to be the first among my friends and family to try something new.	23.8	22.2	38.1* ↑	39.6* ↑
I like to tell others about new brands or technology.	51.8	45.0*	60.2* ↑	62.1* ↑

* $p < .05$ Difference compared to ANES Web Panel uses Fisher's exact test

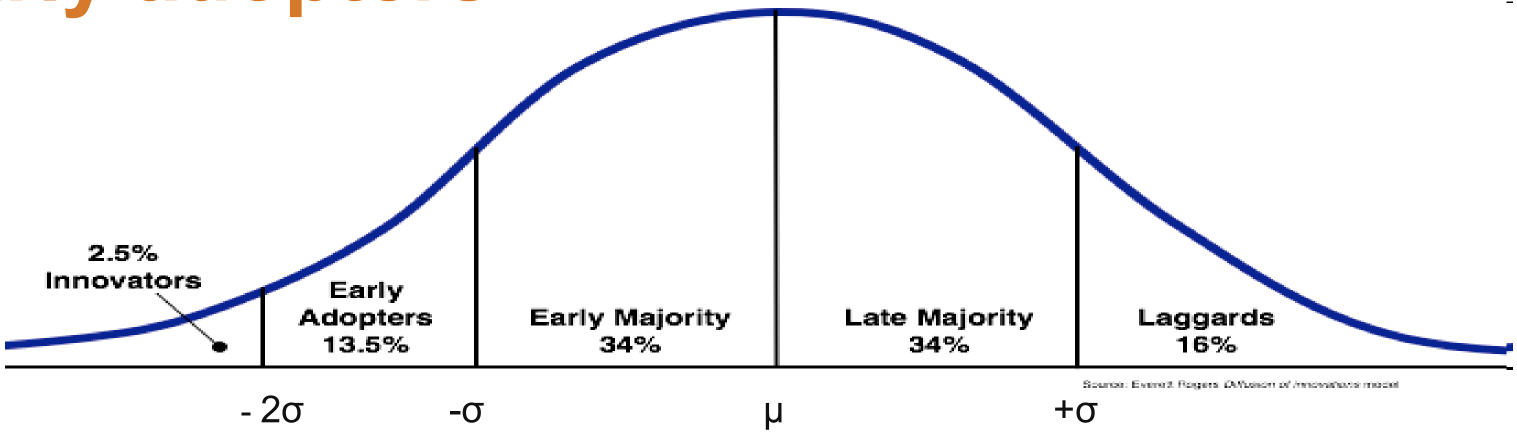
Completion rates: ANES 65.8% ; KN 63.7%; Opt-in A 4.6%; Opt-in B 4.7%

Dennis, J.M., Osborn, L., Semans, K (2009). Comparison Study: Early Adopter Attitudes and Online Behavior in Probability and Non-Probability Web Panels (Accuracy's Impact on Research). Palo Alto, CA: Knowledge Networks.

GfK Knowledge NETWORKS

18

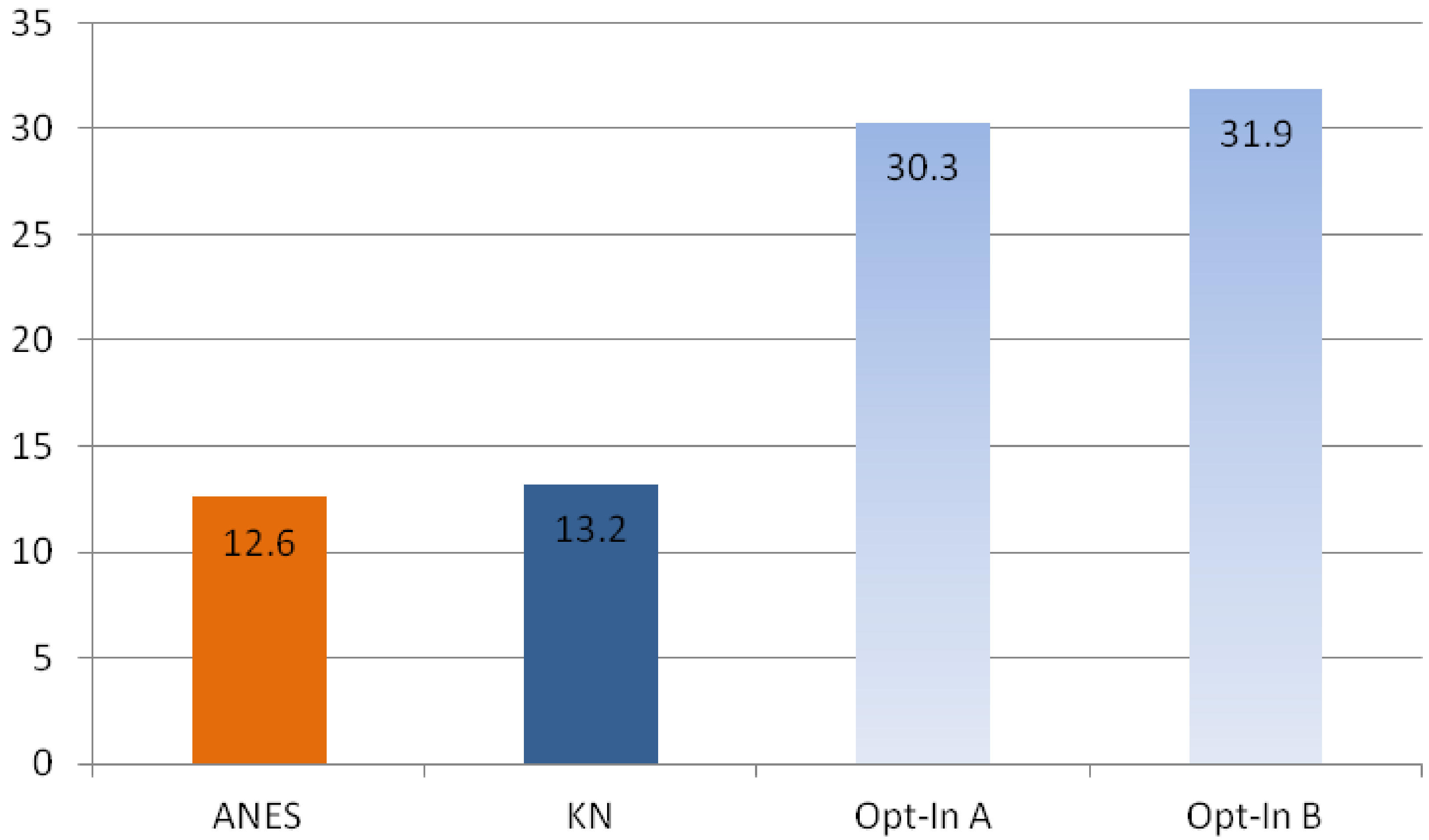
Early adopters



Based on Rogers, M. (2003). Diffusion of Innovations. 5th Edition, New York: Free Press.

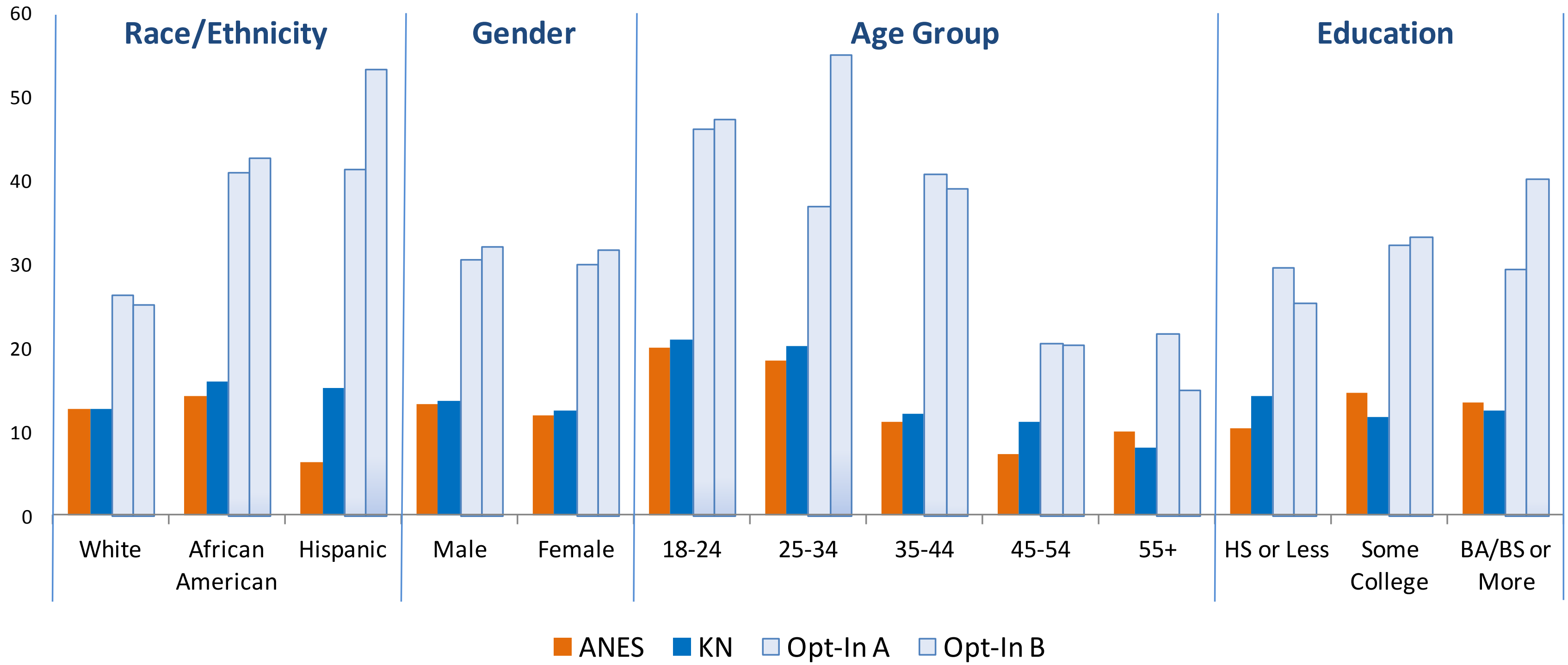
- KN identifies early adopters as respondents with a total EA score that is 1 standard deviation or greater than the estimated population mean.
- Use the full ANES Panel to set the cut-points for all panels.
  - Sum EA question responses for each respondent to calculate a total score.
    - Strongly agree =1, Agree=2, Disagree=3, and Strongly disagree=4
  - Calculate the sample mean (13.6) and sample standard deviation (2.9).

Early adopters by panel (percent)

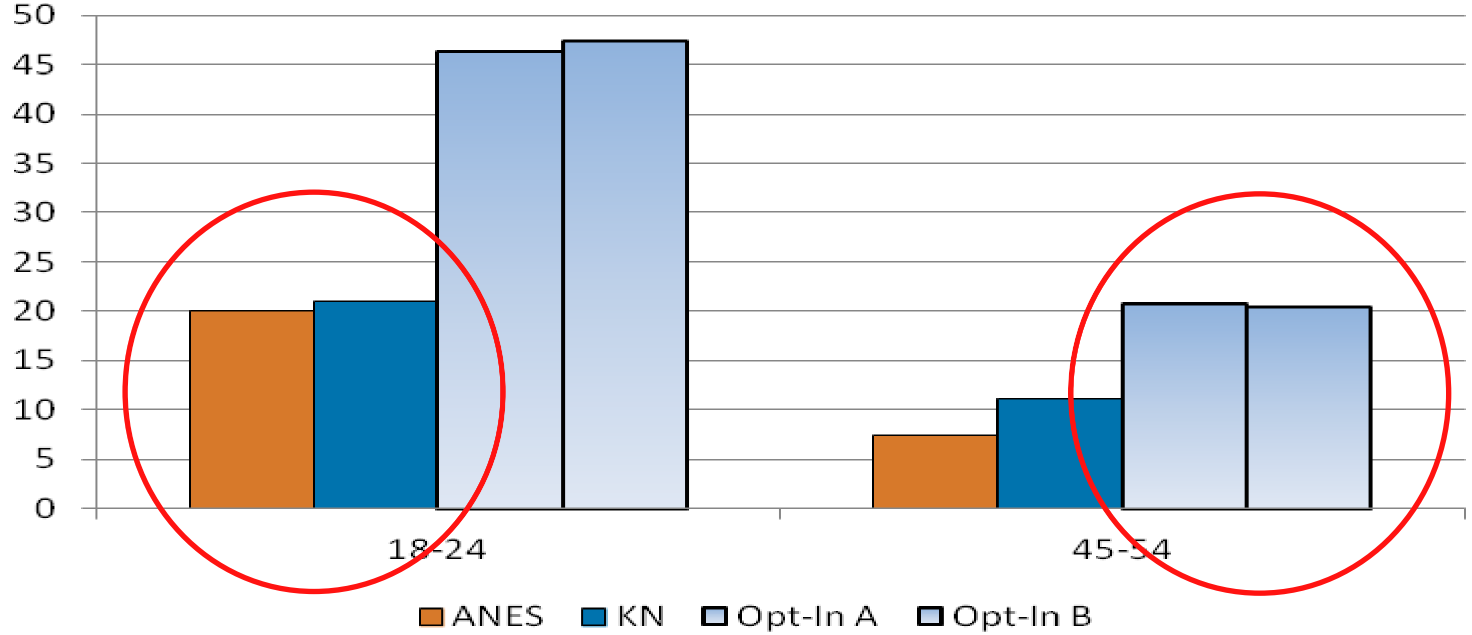


GfK Knowledge NETWORKS

Early adopters by demo group by panel



Early adopters by demo group by panel



- Early adopter attitudes do not always discriminate between probability-based panelists and opt-in panelists when the two samples are comprised of different demographic groups.

GfK Knowledge NETWORKS
22

KnowledgePanel Calibration[SM]

2-step solution and assumptions

1. Supplement probability sample with opt-in panel cases

Use quota sampling with opt-in cases to minimize demo skews/weights

2. Calibrate the combined samples to the probability sample

Assumption A: The probability sample has the most accurate answer

Assumption B: The two samples consist of the same demographic

Assumption C: EA attitudes differentiate the two samples

Assumption D: Weighting on EA attitudes will align the combined samples with the probability sample results

Calibration weighting

Combines data from different sources and uses estimates from one source as “benchmarks” to adjust (calibrate) the data.

- Integrates auxiliary information irrespective of relationship to other variables *(Reuda et al. 2007)*
- Reduction of bias (non-response, coverage, measurement error) *(Kott 2006; Skinner 1999)*
- Efficient for limited time-frames, resources (a lower analyst burden 🙂)
- Can be used for any variable of interest if:
  - differential mode effects are avoided
  - opt-in sample uses quotas to control for demos and impact on weights
  - identified characteristics differentiate opt-in from probability samples

Rueda, M., et al. (2007). Estimation of the distribution function with calibration methods. *J Stat Plan Inference* 137(2): 435–448.

Kott, P. (2006). Using calibration weighting to adjust for nonresponse and coverage errors. *Survey Methodology*, 133–142.

Skinner, C.J. (1999). Calibration weighting and non-sampling errors. *Research in Official Statistics*, *2*, 33-43.

Calibration steps

Step 1 – Weight probability sample

- Weight KnowledgePanel cases only (probability sample) using "standard" demographic/geographic variables
- Use each panel member's base weight (bw_i^{KP}) as the starting weight in a post-stratification raking procedure
- Trim final weights (W_i^{KP}) to control for outliers *(~1st & ~99th %iles)*

$$\sum_{i=1}^{n^{KP}} W_i^{KP} = \sum_{i=0}^{n^{KP}} \left(bw_i^{KP} \times w_i^{KP}\right)$$

where:

bw_i^{KP} = KnowledgePanel member base weight

w_i^{KP} = KnowledgePanel member post-stratification adjustment factor

Use weighted probability sample as benchmark for Step 2

GfK Knowledge NETWORKS

Calibration steps

Step 2 – Combine weighted probability sample with non-probability volunteer opt-in sample

- Use panel member's Step 1 weight (W_i^{KP}) as starting weight
- Set volunteer cases base weight (bw_i^{Vol}) to 1.0 as starting weight
- Weight standard variables to Step 1 benchmarks
- Trim final weights (W_i^{All}) to control for outliers *(~1st & ~99th %iles)*

$$\sum_{i=1}^{N^{All}} W_i^{All} = \sum_{i=1}^{n^{KP}} \left(W_i^{KP} \times w_i^{All}\right) + \sum_{i=1}^{n^{Vol}} \left(bw_i^{Vol} \times w_i^{All}\right)$$

where:

w_i^{All} = All cases blended post-stratification adjustment factor

bw_i^{Vol} = 1.0 for opt-in volunteer panel base weight

This is the weighted "blended" sample for step 3

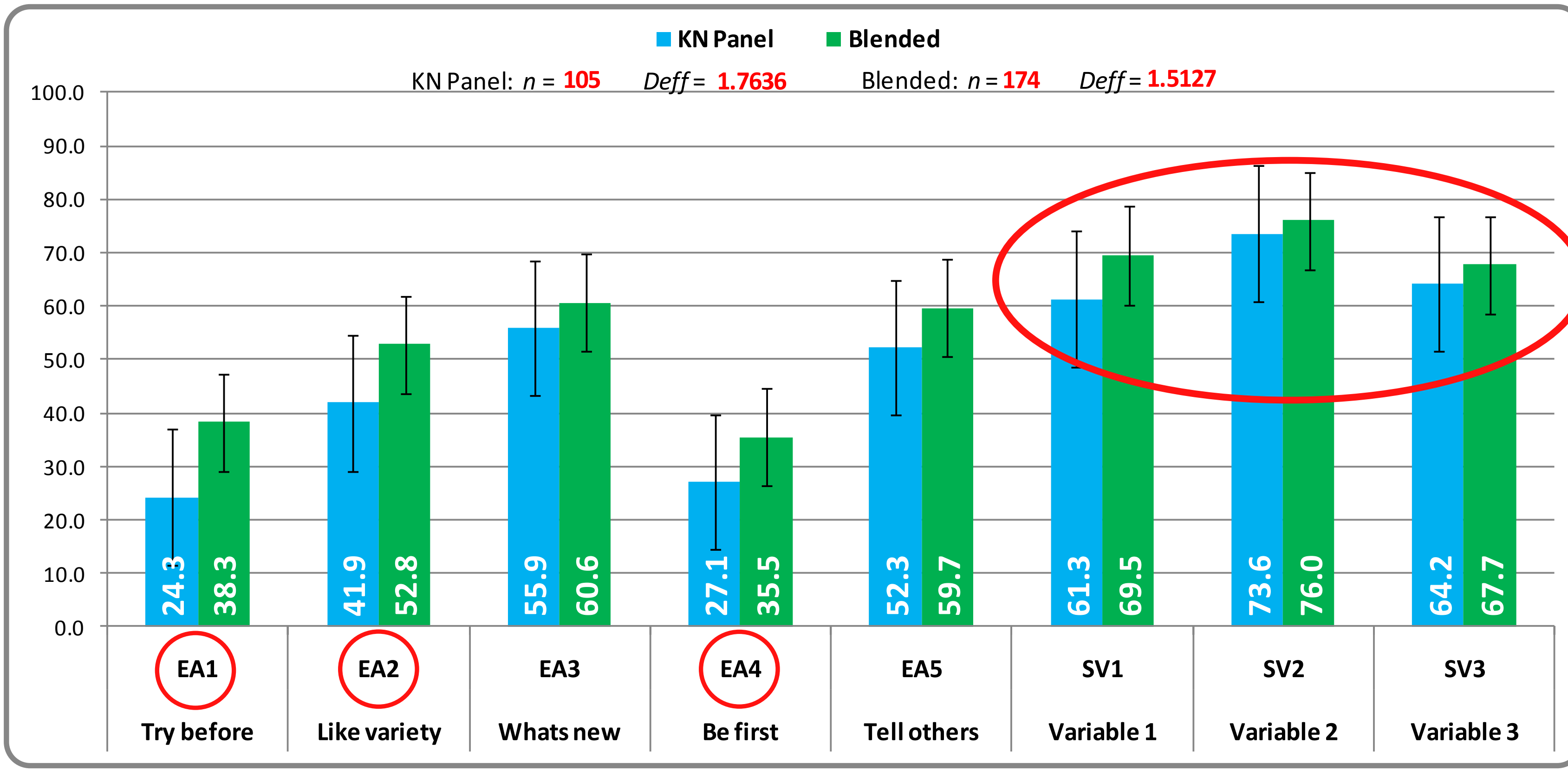
Calibration steps

Step 3 – Evaluate by comparing probability sample (Step 1) to blended sample (Step 2) for:

- ✓ 5 Early Adopter questions (EA1 – EA5)
- ✓ at least 3 study variables (SV1 – SV3)

Results before calibration

Step 4 – Select 1-3 EA Qs as calibration variables for raked reweighting

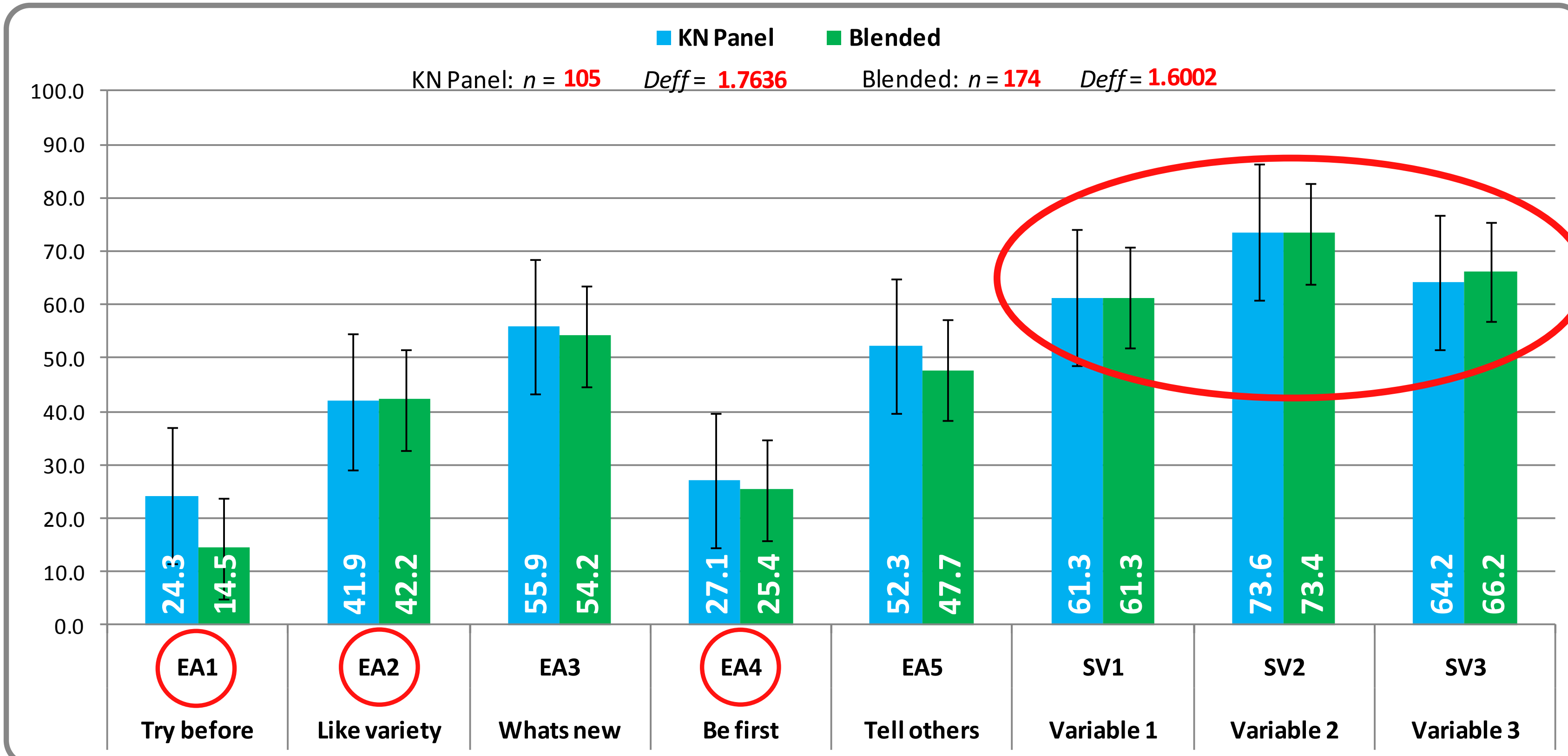


Knowledge NETWORKS

Results after calibration

Evaluate

- Minimize bias introduced from opt-in non-probability sample



Quantitative Analysis

Analysis comparing calibration results

Examined:

1. Weighted probability sample (Reference benchmark)
2. Weighted opt-in sample
3. Blend weighted probability sample with unweighted opt-in sample, then reweight to reference benchmark – no calibration
4. Blend weighted probability sample with unweighted opt-in sample, then reweight to reference benchmark – calibrated

Analysis comparing calibration results

Example 1: Smoking behavior in a mid-west state

Sample

611 probability sample cases
750 opt-in non-probability sample cases
1,361 combined or "blended" cases

Used 13 items from the study questionnaire

Quantitative benchmarks

Examined:

- ✓ Average absolute error in weighted estimates
- ✓ Number of items with an error of 2 percentage points or more
- ✓ Design Effect [$\text{Deff} = \Sigma\, w_i^2 / \Sigma\, w_i$]
- ✓ Average estimated squared bias *(Ghosh-Dastidar et al. 2009)*

$$\hat{\varepsilon}^2 = \max\left(0, (\bar{x}_{Set\ 1} - \bar{x}_{Set\ x})^2 - \frac{s^2_{Set\ 1}}{n_{Set\ 1} - 1} - \frac{s^2_{Set\ x}}{n_{Set\ x} - 1}\right)$$

- ✓ *Average estimated Mean Squared Error*

$$MSE_{Set\ x} = \hat{\varepsilon}^2 + \frac{s^2_{Set\ x}}{n_{Set\ x} - 1}$$

Ghosh-Dastidar, B., Elliott, M. N., Haviland, A. M., & Karoly, L. A. (2009). Composite Estimates from Incomplete and Complete Frames for Minimum-MSE Estimation in a Rare Population: An Application to Families with Young Children. Public Opinion Quarterly, 73 (4), 761-784.

Analysis comparing calibration results
Example 1: Smoking behavior in a mid-west state

	Weighted probability sample (Reference)	Weighted opt-in sample	Two samples blended, re-weighted No calibration	Two samples blended, re-weighted Calibrated *
Number of cases	611	750	1,361	1,361
Average Absolute Error	--	5.3%	2.3%	1.3%
No. of items with error of 2 or more percentage points	--	12	7	3
Deff	1.872	3.480	2.155	2.095
Average Est. Squared Bias	--	25.579	2.056	0.064
Average Est. MSE	3.937	28.741	3.816	1.826

* Calibrated using EA1, EA3 and EA5

Analysis comparing calibration results
Example 2: Environmental attitudes in a coastal state

	Weighted probability sample (Reference)	Weighted opt-in sample	Two samples blended, re-weighted <u>No calibration</u>	Two samples blended, re-weighted <u>Calibrated</u> *
Number of cases	1,280	767	2,047	2,047
Average Absolute Error	--	9.4%	3.5%	2.6%
No. of items with error of 2 or more percentage points	--	11	10	8
Deff	2.369	1.734	2.162	2.190
Average Est. Squared Bias	--	103.425	13.266	6.213
Average Est. MSE	1.807	106.389	14.402	7.347

* Calibrated using EA1, EA2 and EA3

Analysis comparing calibration results
Example 3: Chain restaurant usage among Hispanics

	Weighted probability sample (Reference)	Weighted opt-in sample	Two samples blended, re-weighted No calibration	Two samples blended, re-weighted Calibrated *
Number of cases	506	251	811	811
Average Absolute Error	--	10.1%	3.2%	2.0%
No. of items with error of 2 or more percentage points	--	10	7	5
Deff	2.406	1.738	2.152	2.083
Average Est. Squared Bias	--	142.845	10.570	4.175
Average Est. MSE	4.259	152.310	13.548	7.153

* Calibrated using EA1, EA2 and EA4

Analysis comparing calibration results
Example 4: Holiday party shopping among Hispanics

	Weighted probability sample (Reference)	Weighted opt-in sample	Two samples blended, re-weighted No calibration	Two samples blended, re-weighted Calibrated *
Number of cases	532	268	800	800
Average Absolute Error	--	15.9%	5.6%	4.5%
No. of items with error of 2 or more percentage points	--	12	11	11
Deff	2.267	1.881	2.032	2.080
Average Est. Squared Bias	--	275.285	29.873	17.078
Average Est. MSE	4.062	284.001	32.732	19.904

* Calibrated using EA2, EA3 and EA4

Conclusions

- Calibrating non-probability sample with probability sample using early adopter questions minimizes bias in the resulting larger combined sample
- The KnowledgePanel Calibration^{SM} technique can deliver larger sample sizes when the preferred probability sample source is limited due to panel size
- Process serves short timelines, rapid data turnaround

Future research

- Identify additional characteristics and attitudes that generally distinguish between probability-based panelists and opt-in panelists and can be used for calibration
- Continue to evaluate our calibration approach across survey topics and populations
- Continued research is necessary to better understand the underlying statistical implications

Q & A

Contact: Charles DiSogra
cdisogra@knowledgenetworks.com

Thank you!