

Do We Hear Different Voices?

Investigating the differences between Internet and non-Internet users on attitudes and behaviors

Chan Zhang, University of Michigan

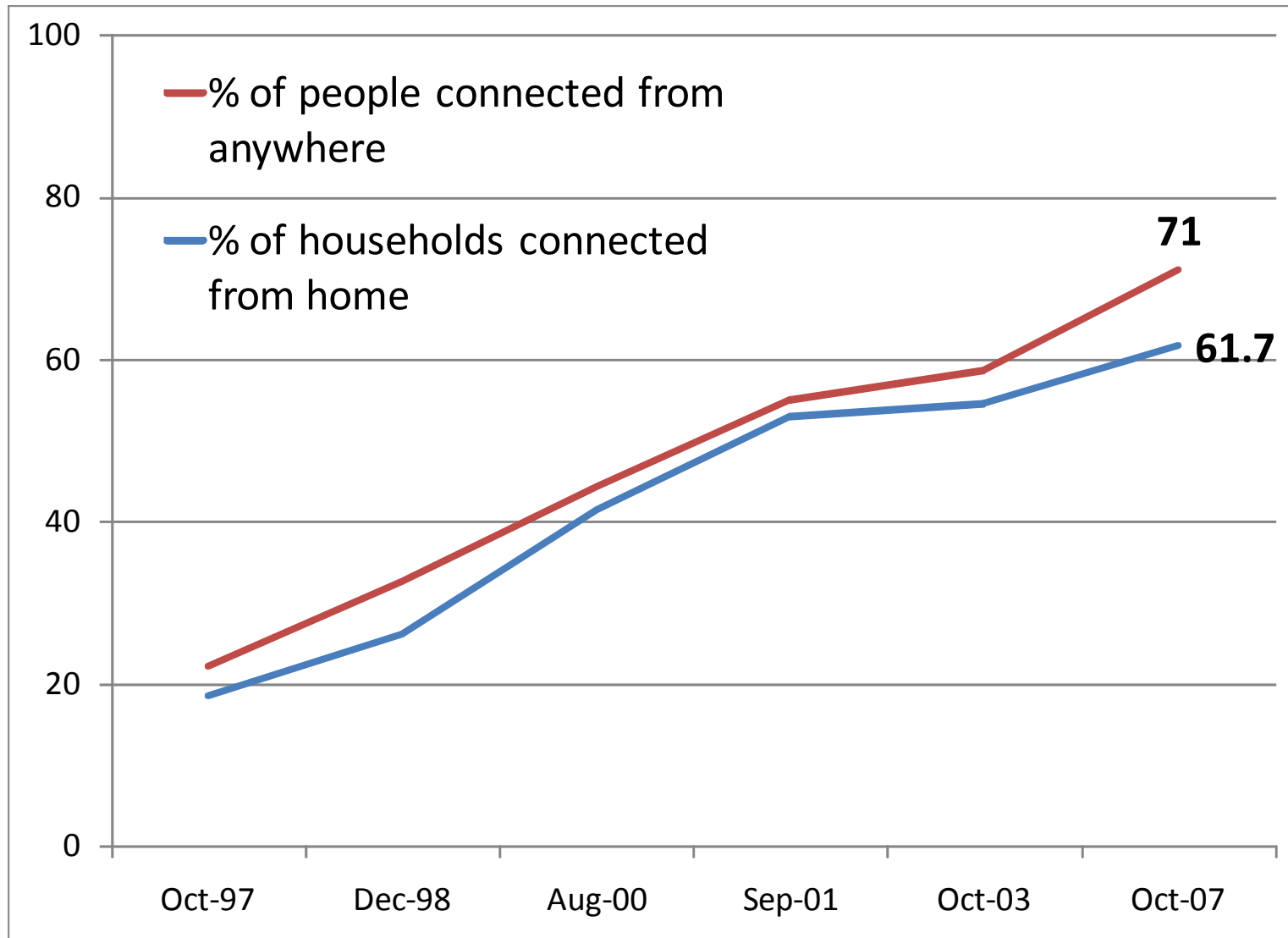
Mario Callegaro, Melanie Thomas & Charles DiSogra

Knowledge Networks, Inc.

Prepared for:

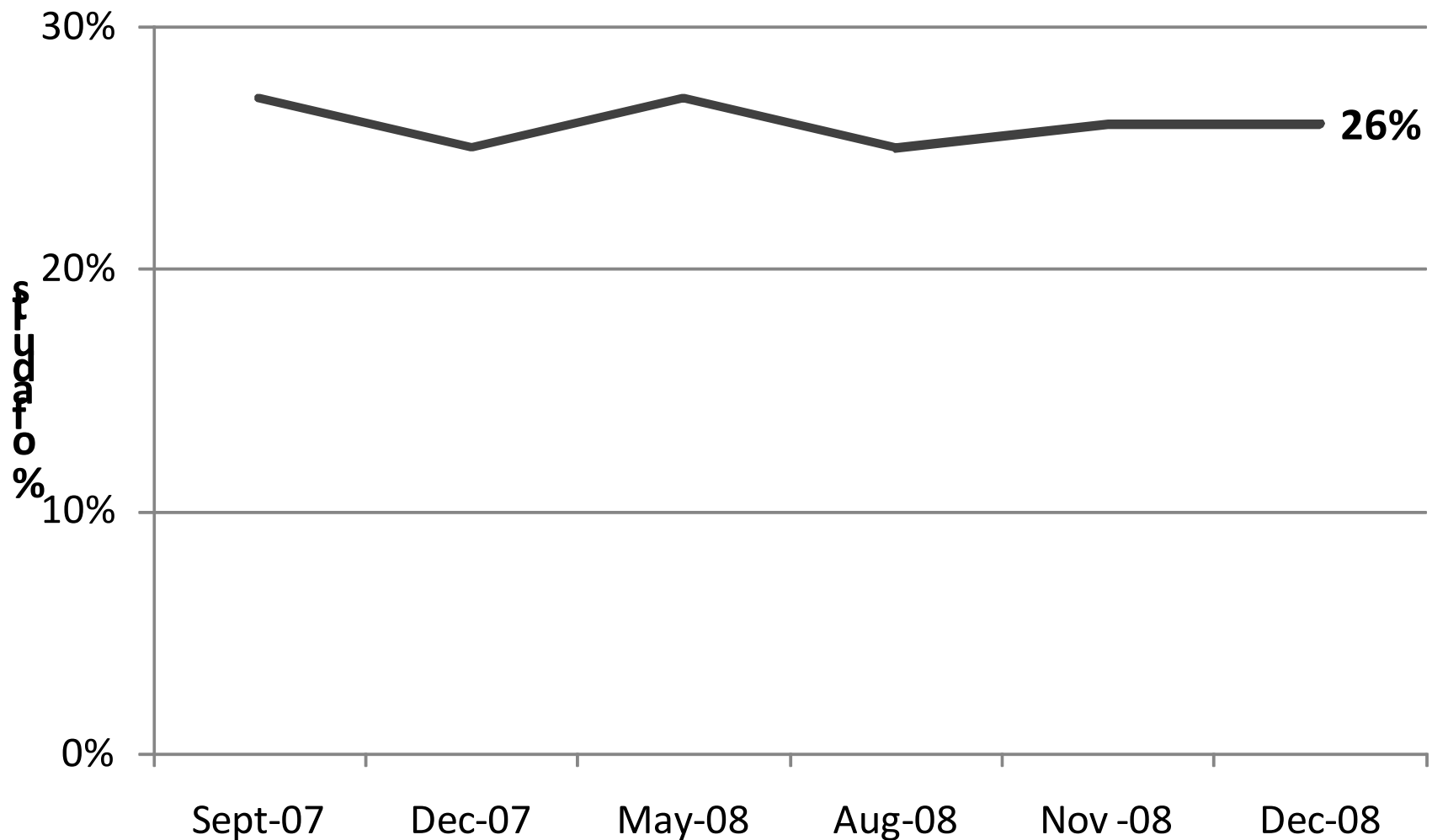
2009 AAPOR Conference

Still 38.3% of Households are not Connected (CPS Oct. 2007 Supplement)



Is the Digital Divide Closing? (More Recent Data, Adults)

- Percent of “truly disconnected” (*have never used the Internet or email*) has remained stable (data from Pew 2009)



Previous Research I

- The demographic differences between Internet and non-Internet users remain (e.g., age, education, ethnicity, income, presence of kids in the HH...)
- When using probability based samples, differences between Internet and non-Internet users still remain after weighting or controlling for demographics correlates on a variety of variables (e.g., Martin & Robinson, 2005; Rookey, Hanway and Dillman, 2008), with some exceptions (e.g., Dever, Rafferty and Valliant, 2008), although size of weights not shown

Previous Research II

- When non probability based samples are used (e.g., opt-in online panels), weighting approaches to compensate for non-Internet users fail
- “Internet users that belong to Access Panels are different from all internet users and frequent users” (Piekarski et al, AAPOR 2008)

Divide in Access and Noncoverage Error of Web Surveys

A major concern with Web surveys is noncoverage error:

- If non-Internet users/households are excluded
- If their responses differ from the Internet users/households

Approaches to address the concern:

- Post-survey adjustments: the key is to find the adjustment variables that are correlated to both the Internet use and the variable of interest
 - Concern: limited choice on the adjustment variable, as not much information is available for the study population

Divide in Access and Noncoverage Error of Web Surveys (cont'd)

Approaches to address the concern (cont'd):

➤ Including non-Internet users:

- Mixed mode: non-Internet users are given options to respond in methods other than the Web
 - Concerns: mode effects; differences in turnover time, limitations of the mode (e.g., mail questionnaire cannot deliver all stimuli and support complex logic)
- Web only: non-Internet users are provided equipment so that they can respond online
 - Concern: may have difficulties recruiting those who are “truly offline”

Study Purpose

Research question: To what extent can the demographic differences explain the differences in responses between the Internet and non-Internet HH in the Web panel?

Use data from KnowledgePanel®

- Panel recruited by RDD (at the time this study was done) and now via ABS frame
- Non-Internet HH used to be given Web TVs and now laptops

Data on the panel members were from Core Adult Profile Survey late 2007 to April 2008

Method

Logistic Regression Analysis (categories combined, if necessary)

Differences between the Internet and non-Internet HH are measured by odds ratios

Four models with increasing control on the demographic characteristics

- Univariate analysis: *access status*
- Basic model:
access status + (income, age, education, race)
- Intermediate model:
Basic Model + (gender, region, metropolitan status)
- Full model:
Intermediate Model + (employment status, marital status + kids in HH + home ownership)

Results

Different patterns are found of how the odds ratio changes among the four models

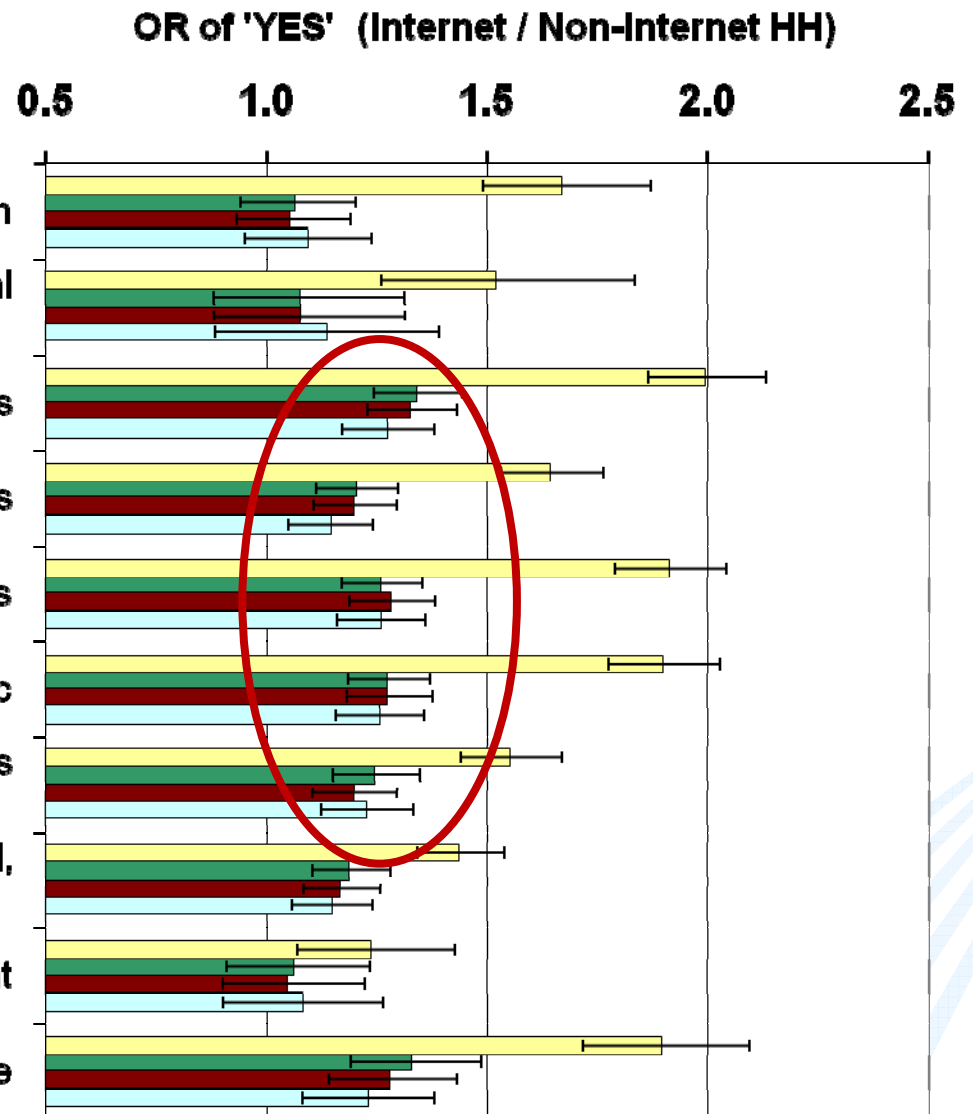
- **Pattern A**: Controlling demographic characteristics reduces differences between the two groups. But adding more demographic variables does not help much

Pattern A: Example

In the past 12 months, have you ...?



- Given money to an environmental organization
- Contributed time or expertise to an environmental group
- Recycled newspapers or other papers
- Recycled cans
- Recycled glass
- Recycled plastic
- Purchased a product specifically because it was made from recycled materials
- Taken steps to reduce the use of energy (oil, gasoline, electricity)
- Taken other actions to improve the environment
- Any of these



Differences in Responses: Results

Different patterns are found of how the odds ratio changes among the four models

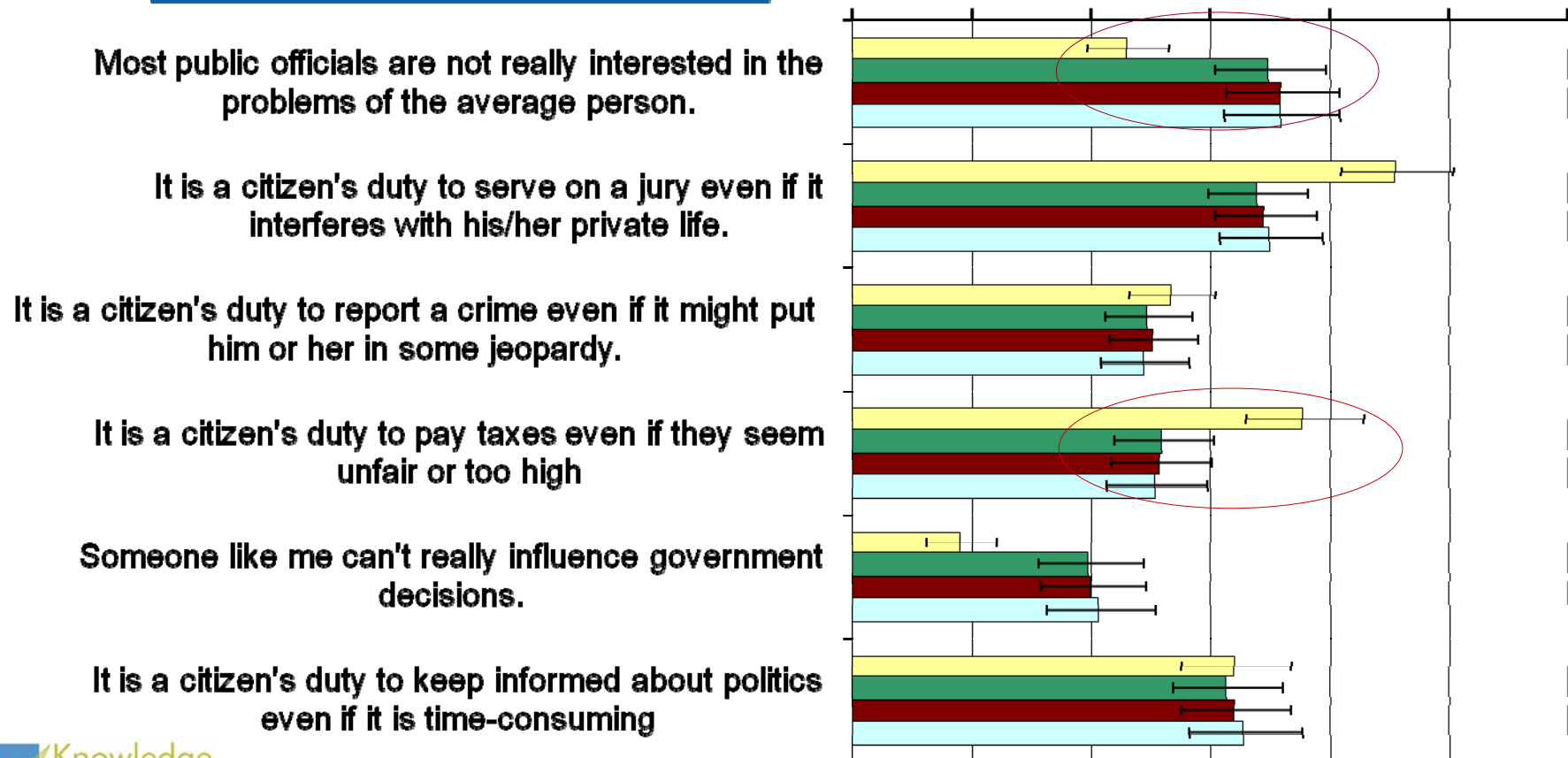
- **Pattern B**: Controlling demographic variables may flip the difference observed in simple comparison (i.e. univariate analysis)

Pattern B: Example

Do you agree or disagree with the following statements about political affairs in this country?



OR of 'strongly agree' (Internet / Non-Internet HH)
0.4 0.6 0.8 1.0 1.2 1.4 1.6



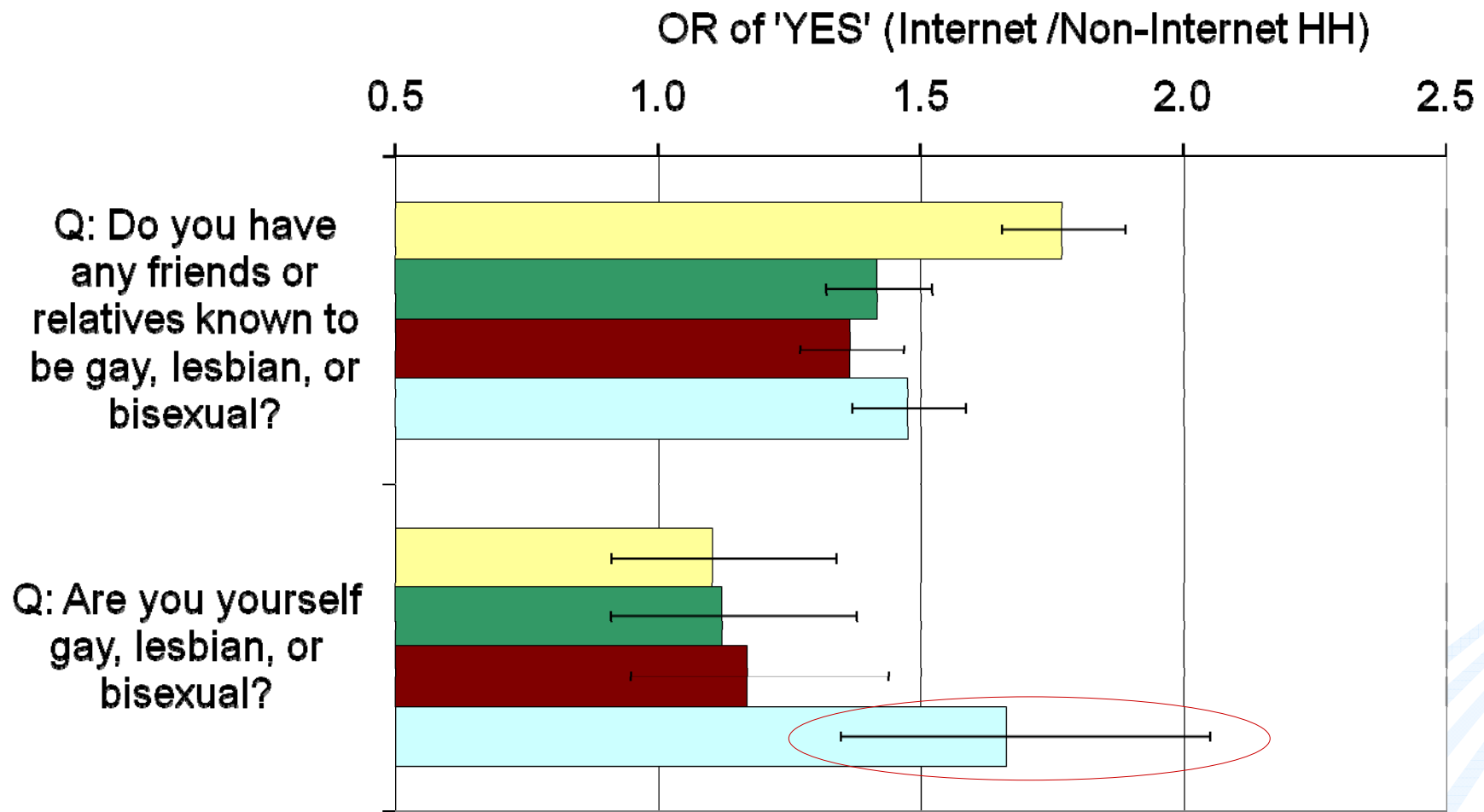
Differences in Responses: Results

Different patterns are found of how the odds ratio changes among the four models

- **Pattern C**: Demographic variables identifies the differences disguised in the simple comparison

Pattern C: Example

■ univariate ■ basic model ■ medium model ■ full model



Conclusions

This study suggests that

- Differences in survey responses between Internet HH and non-Internet HH may still exist, even when many demographic characteristics are used in the adjustment
- Adjustment on only 4 demographic variables (race, education, income, age) can be as good (or as bad) as that based on 12 variables
- It is difficult to predict what variables should be used to adjust in advance, and in many cases, no combination compensates for the differences