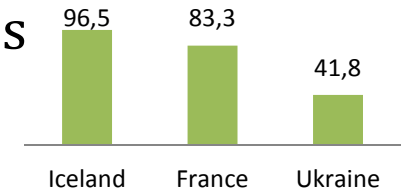
What is the gain in a probability-based online panel of providing Internet access to sampling units who previously had no access?

Melanie Revilla[1], Anne Cornilleau[2], Anne-Sophie Cousteaux[2], Stéphane Legleye[3], Pablo de Pedraza[4]

[1] RECSM - Universitat Pompeu Fabra
[2] Centre de données sociopolitiques (Sciences Po / CNRS)
[3] National Institute for Demographic Studies (INED); Inserm; University Paris-Sud and University Paris Descartes; University Paris XI
[4] University of Amsterdam, AIAS, University of Salamanca

Coverage problem

- Internet coverage very high in European countries
  - But differences across countries



- Still, a proportion of the population without access
- People with Internet access differ from those without it
  - In terms of age, income and education
  - Weighting not enough to correct for bias
- In order to use Internet to survey the general population, different possibilities
  - Mixed-mode surveys
  - Provide Internet access to non-Internet units
    - Knowledge panel (USA), LISS panel (NL)
    - GIP (Germany), ELIPSS (France)

Non participation and non response problem

- Providing Internet access can solve the coverage problem
- But problem of representativeness still possible
  - If the non-Internet unit refuse to participate even if provided with Internet access
  - Some empirical evidence supports this hypothesis
  - Leenheer & Scherpenzeel (2013): recruitment rates for the non-Internet households (35%) were much lower than those for households with Internet access (84%)
- Getting in the panel ≠ answering each single survey
  - Non response in each survey may also be correlated with previous Internet access
  - Attrition too
- Necessary to study representativeness even for probability-based panels providing Internet access
  - Done in previous research for Knowledge panel, LISS...
  - Overall, representativeness quite good

Additional gain

- Providing Internet access increases a lot the cost
  - Material, installation, but also preparation, recruitment procedure, etc
- If non coverage is quite low in a country, and non participation + non response is higher for non-Internet units, we can wonder:

What is the exact gain of providing the access?
Is it worth the extra costs?
Would the representativeness of the probability-based panels be (much) lower if not providing the equipment to non-internet units?

- Previous research on this
  - Leenheer & Scherpenzeel (2013): "The research question is whether providing non-Internet households with a means of Internet access, which requires a substantial investment, significantly improves the quality of an Internet panel"
- Our study:
  - Different country: France
  - Different panel: ELIPSS

ELIPSS panel

- Part of the *Data, Infrastructure, Methods of Investigation in the Social Sciences and Humanities* (DIME-SHS) project, led by Sciences Po
- Probability-based Internet panel, inviting researchers to submit survey projects during calls for proposals
- A scientific committee evaluates the applications on the project's research purpose
- Commercial use excluded
- Pilot started in 2012 and consisted of 1,039 panel members

ELIPSS panel

- Differs from others
  - Tablets and a 3G connection are offered to all panel members, and not just to those who did not previously have Internet access
  - All ELIPSS panellists use the same device and browser to answer questionnaires
  - No additional monthly incentives
- Can create differences from previous studies

How to define Internet access?

- To study the gain in a probability-based online panel of providing an Internet connection to units who previously had no Internet access

- We need to define “no previous Internet access”!!

What is it to have Internet access?

How to define Internet access?

What is it to have Internet access?

How to define Internet access?

Training-Division

31.	Training within my division helps me do my job better.	Strongly Disagree	1	2	3	4	5	6	7	Strongly Agree	N/A
32.	Resources are available to me (e.g., manuals, job tools) that help me do my job better.	Strongly Disagree	1	2	3	4	5	6	7	Strongly Agree	N/A

Communication-Organization

33.	I regularly receive information that lets me know what's going on in the organization.	Strongly Disagree	1	2	3	4	5	6	7	Strongly Agree	N/A
34.	Administrators share information across the organization.	Strongly Disagree	1	2	3	4	5	6	7	Strongly Agree	N/A
35.	In this organization, it's easy to obtain	Strongly Disagree	1	2	3	4	5	6	7	Strongly Agree	N/A
	ually looks for e get information.	Strongly Disagree	1	2	3	4	5	6	7	Strongly Agree	N/A
	anyone in the	Strongly Disagree	1	2	3	4	5	6	7	Strongly Agree	N/A

I would like...	Strongly disagree 1	2	Neither agree nor disagree 3	4	Strongly agree 5
(a) to earn special avatar gear within a game (e.g., a special hat) just like I can earn achievements	○	○	○	○	●
(b) my avatars to be better integrated into the messages I send to friends	●	○	○	○	○
(c) to make a set of "family" avatars. For example, having better representations of kids or pets	○	○	●	○	○
(d) the characteristics of my avatar to affect my in-game skills or options (e.g., an athletic looking avatar is faster)	○	○	○	●	○
(e) to have themed avatar gear (e.g., a St. Patrick's Day theme)	○	○	○	○	●
(f) to make my avatar look like a celebrity, game character or other famous personality	●	○	○	○	○
(g) an automatic way to make my avatar look like me	○	○	○	●	○
(h) my avatar to be themed to the game I'm playing while online (e.g., wearing a pirate suit while playing a pirate game)	○	○	○	○	●
(i) to play more games as my avatar, within the game	○	○	○	○	●
(j) to be able to interact with others via avatars in a "virtual" social setting	○	○	○	○	●
(k) to be able to create and control simple, unique animations with my avatars to share with others	○	○	○	●	○
(l) to be able to customize my avatar's features (e.g., hair, face, body-type) in a more granular way	○	○	●	○	○
(m) to be able to create my own cloths, etc. for my avatar	○	○	●	○	○
(n) to use my avatar outside of my Xbox. For example on Facebook or MySpace	○	○	●	○	○
(o) to have my avatar express my current mood (happy, sad, crazy, etc.)	○	○	○	●	○
(p) to be able to see my friends' avatars in crowds or elsewhere in games, even when they are not online	○	○	○	○	●
(q) to have different avatars for different games / situations / days of the week	○	●	○	○	○
(r) to be able to collect other avatar related goodies such as pets or cars	○	○	○	○	●
(s) to purchase items for my avatar	●	○	○	○	○

How to define Internet access in the frame of a web panel?

Information we use to define Internet access

- Two sources
  - Recruitment: one person in the household was asked questions about the household computer and Internet equipment
  - First survey (December 2012 - March 2013) about Internet access and digital practices, before joining the ELIPSS panel

- Information not always the same for a given unit in both sources
  - Respondent during the recruitment phase may be different from the panellist
  - Formulations of the questions slightly different
  - Possible that the situation changed
  - Measurement errors
  - We crossed the two sources of information
  - For this presentation, we focus on the *Priority Survey* definition
    - Use information from the survey as main source
    - Complete with recruitment for whose who did not answer the survey
    - Advantage: information comes from the panellist, more recent information
    - Other ways of crossing information lead to differences in results

Number of additional panelists

- As a reference: 14.5% of the 18-75 year-olds do not have Internet in metropolitan France (2013 ICT survey)
- Proportion of non-Internet units

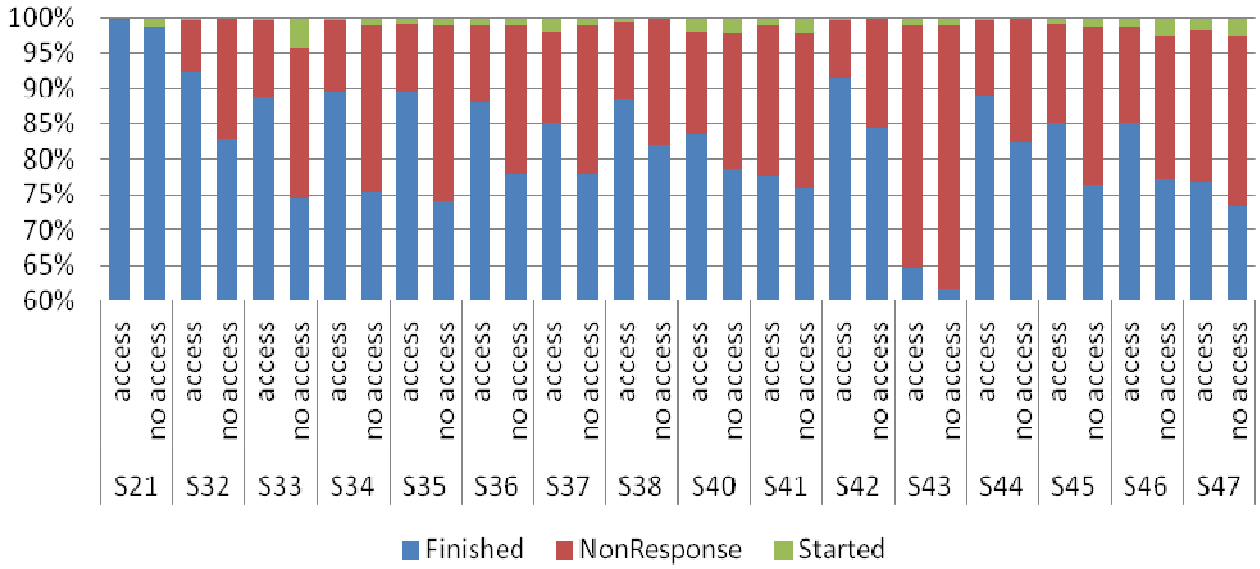
Accept	Internet	Freq.	%
Household level (n=1334)	Yes	1176	88.2
	No	158	**11.8**
Sign agreement (n=1036)	Yes	938	90.5
	No	98	**9.5**
Answer 1st survey (n=939)	Yes	863	91.9
	No	76	**8.1**

Decreasing each step

- Out of the 76 answering the 1st survey
  - 31 still connecting to Internet everyday or almost ➔ Coverage ≠ usage
  - 25 connecting from home… ➔ difficulty defining access

Participation in surveys

- Participation of invited panelists in the different surveys, differentiating between Internet ("access") and non-Internet ("no access") units



- High participation overall
- But proportions who finished the survey systematically lower for the non-Internet units
  - Not more loyal (≠ LISS, because incentives also ≠ ?)
- No trend over time

Representativeness

- Definition: a group G is representative with respect to variable *i* (e.g. gender) if the distribution of the variable *i* in G is similar to the one in the population of interest

- Compare characteristics of the whole group of panellists and respondents of selected surveys with the population of interest
  - On primary and secondary socio-demographics variables

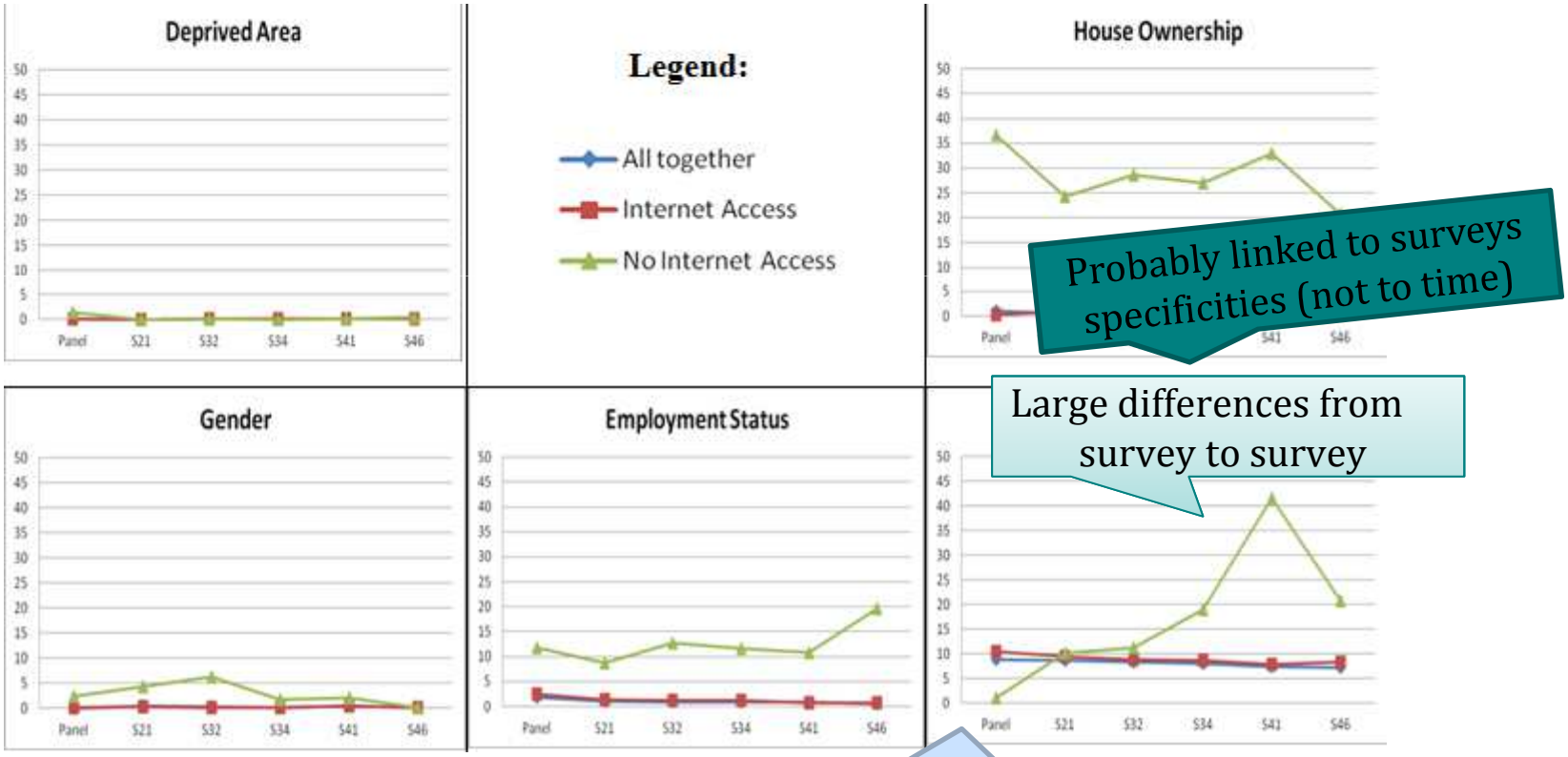
- Population of interest approximated with 2012 French LFS (selecting 18-75 year olds)

- Compute Chi^2 distance between each group and the target population

$$d^2 = \sum_{i=1}^{k} \frac{(Observed_i - Theoric_i)^2}{Theoric_i}$$

Distance Chi^2

Except “deprived” area, the Chi^2 distance is different in the group without Internet access

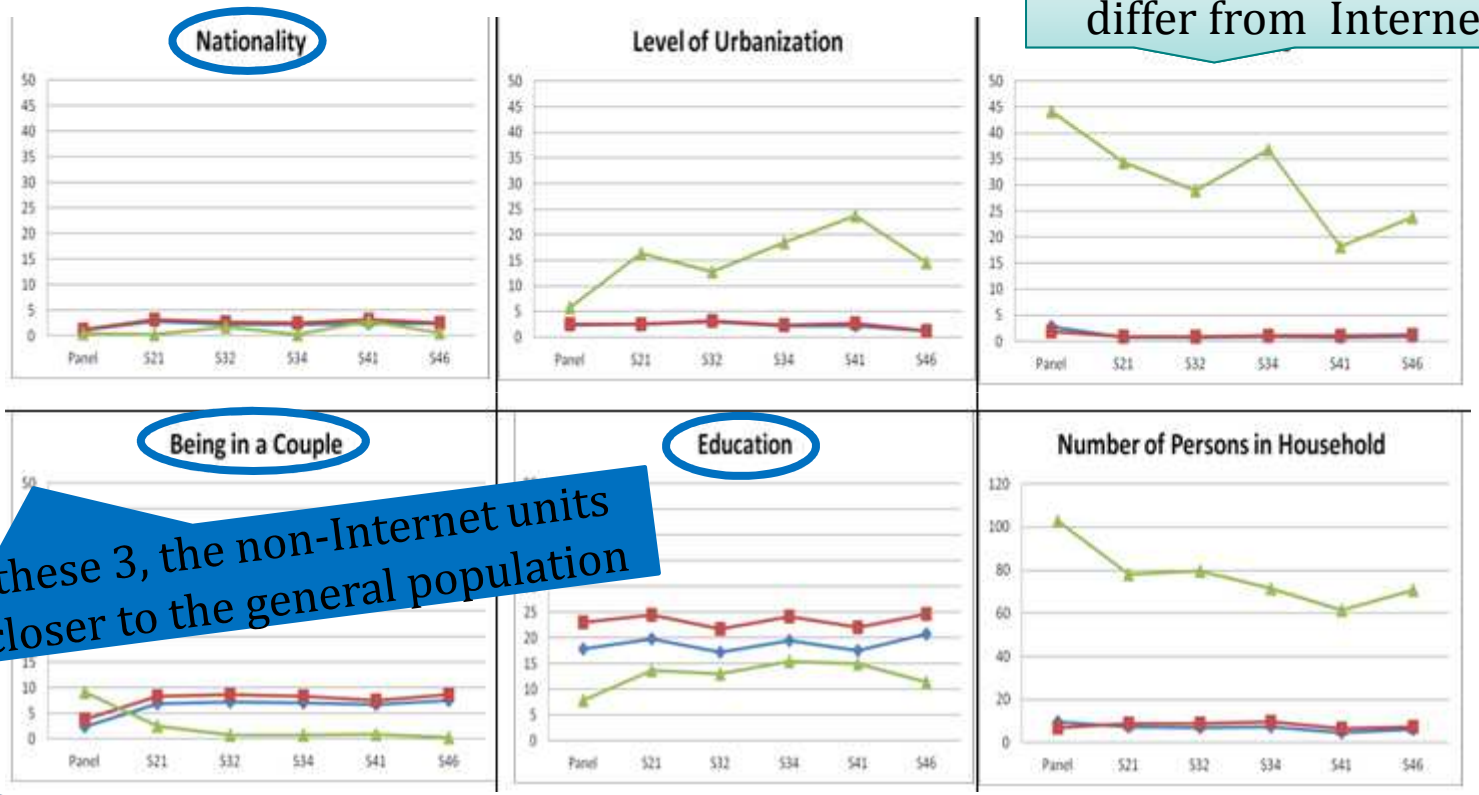


Probably linked to surveys specificities (not to time)

Large differences from survey to survey

Some differences between panel and surveys

Distance Chi^2



Clear that non-Internet units differ from Internet units

For these 3, the non-Internet units are closer to the general population

Including non-Internet units improves representativeness

Conclusions

- Difficulty defining Internet access
  - Within “non-Internet” units, 41% connected everyday and 33% connected from home
  - Access ≠ usage ≠ using it for answering surveys
- Low number of extra units
  - Linked to high coverage
  - Small size of the pilot panel
- These units are participating less in the different surveys
  - Lower participation
  - In fact 2 groups: one very loyal and one not

Conclusions

- But these units are different on most socio-demographic aspects

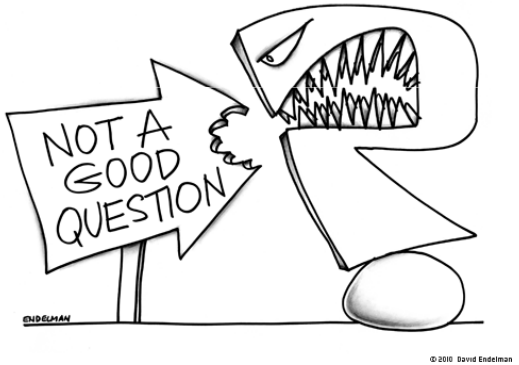
- Non-Internet group more different from the general population for a majority of variables
  - gender, employment status, level of urbanization, home ownership, age, marital status, household size

- For a few key variables, non-Internet units are closer to the general population
  - Education, being in a couple, (nationality)
  - For these variables, complete sample closer to general population than sample of Internet units only
  - Improves representativeness

Conclusions

- So should future panels provide Internet access?
  - If they can afford it, it can improve a little the representativeness
  - Defining who is considered a non-Internet unit would be crucial
  - However, it seems reasonable to think about panels which would be probability based but not provide access to units without Internet
- Decision also depends
  - on the Internet coverage in the country of interest
  - on the size of the panel
    - Fixed costs for including non-Internet units are high, so if the panel is larger it may be more worth it to face them

Thank you for your attention!



If you want to know more…

Revilla, M., Cornilleau, A., Cousteaux, A.S., Legleye, S., and P. de Pedraza (forthcoming). "What is the gain in a probability-based online panel to provide Internet access to sampling units that did not have access before?" *Social Science Computer Review.* Published online first in June 2015, DOI: 10.1177/0894439315590206

Email:

melanie.revilla@upf.edu