Navigating Challenges Implementing In-the-Moment Surveys with Metered and Geolocation Data

WEB DATA OPP Workshop

2024, March 18-19

CARLOS OCHOA | Research and Expertise Centre for Survey Methodology (RECSM – UPF)

Thanks to Melanie Revilla for her guidance and invaluable feedback throughout this research.

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No849165), PI: Melanie Revilla.

Memory recall errors

"A central tool of social science research (…) is asking people questions about what happened. Because of the critical role of retrospective reports, a major source of error in social science data is ***memory errors****."*

Roger Tourangeu (2000)

Memory recall errors

Factors increasing the chances of suffering memory errors:

+ Many events of the same category (e.g., supermarket visits)
+ Low distinctiveness
+ Low emotional impact
+ Short duration
+ Non-rehearsal (time spent thinking or talking about the event).

nothing we can do to attenuate the effect

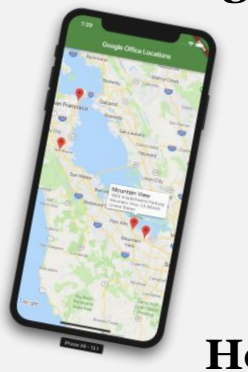
+TIME!

Two potential strategies

Avoid asking (passively collected data)

Modern technological developments have expanded the opportunities for observing behaviors.

Where we go



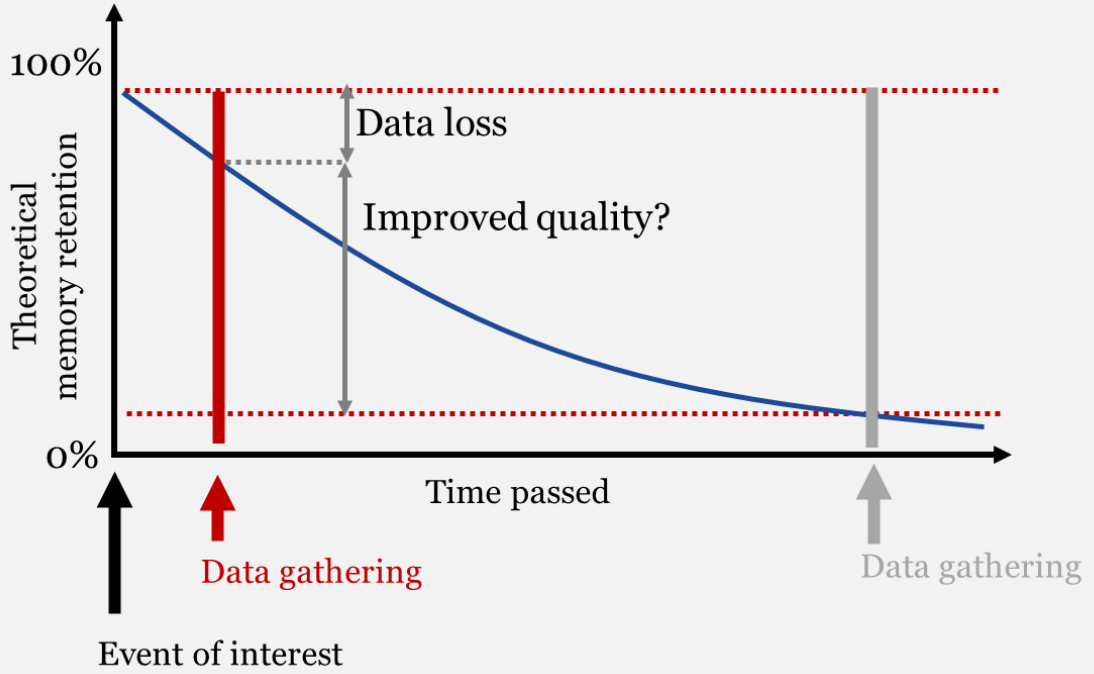
Our SM interaction

What we listen to

How we browse

Surveying participants earlier

Reducing the time gap between the event of interest and data collection = decrease in memory loss.



Two potential strategies

Avoid asking (passively collected data)

Modern technological developments have expanded the opportunities for observing behaviors.

PROS

-Immune to memory errors
-Granularity
-Low burden on participants

CONS

- Unable to gather certain objective data
- Unable to gather subjective data (e.g., motivations)
- Vulnerable to other errors often overlooked (Bosch & Revilla, 2022).

Surveying participants earlier

Reducing the time gap between the event of interest and data collection = decrease in memory loss.

PROS

-As flexible as any survey (all type of data)
-No technology involved

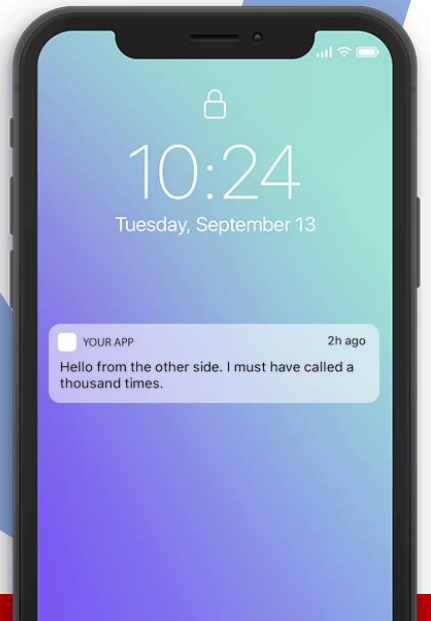
CONS

-Frequent surveying to detect events by chance? (Coincidental surveys; Lamas, 2005)
-Unfeasible in practice

In-the-moment (ITM) surveys: the best of both worlds…

1 Detection of an event of interest using passively collected data

2 Push notification + email invitation "in the moment"



3 Online survey to gather additional objetive data + subjective data

Better data?

New data?

Willingness to participate?

Two experiences within the Web Data Opp project

1 ITM survey triggered by metered data
Completed in 2023

- **Goal**: understanding how people apply for jobs online.
- **Passive data**: metered data
- **Triggering event**: applications at job search websites (e.g., linkedin.com/jobs)
- **Source**: Metered and Survey Panel in Spain (Netquest)
- **Method**: 2 samples of 200 participants:
  - Conventional survey, 160 non-metered and 40 metered participants
  - ITM survey, 200 metered participants.
- **Fieldwork**: from March 10th to October 3rd, 2023.

2 ITM survey triggered by geoloc data
On going

- **Goal**: understanding sunbathing habits at outdoor swimming areas.
- **Passive data**: geolocation data.
- **Triggering event**: entering a geolocation area identified as a beach.
- **Source**: Survey Panel in Spain (Netquest)
- **Method**: 2 samples of ≈400 participants:
  - Conventional survey
  - ITM survey.
- **Fieldwork**: Planned from June 1st to September 31st, 2024.

Setting up ITM projects

Specific tasks compared to conventional surveys

Setting up ITM projects

Specific tasks compared to conventional surveys

Metered data

ITM surveys triggered by metered data (specific tasks)

1 Questionnaire design

FALSE NEGATIVES

Non-detection:
- Applications from non-metered devices
- Temporarily pausing the meter
- Non-detectable applications (see later)

Assuming that this a sample of applications

Researching selection bias? (e.g., surveys asking who uses non-metered devices)

FALSE POSITIVES

False detections:
- Shared-devices = applications from non-participants.
- Risk of revealing job searches from 3rd parties to the participant during questioning.

Filter questions

1. Have you performed any of these activities?
2. Can your confirm that you have just applied for this job?

ITM surveys triggered by metered data (specific tasks)

1 **Questionnaire design**

2 **Elaborating a complete list of job search websites.**

	Website
#1	Linkedin.com/jobs
#2	infojobs.net/
#3	ticjob.es/
#4	es.indeed.com/
#5	es.jooble.org/
#6	infoempleo.com/
#7	Jobtoday.com/es
#8	Insertia.net
#9	Tecnoempleo.com
#10	monster.com/ (.es)
#11	Randstad.es
#12	Adecco
#13	Primerempleo.com
#14	Trabajos.com
#15	Jobatus.es

- General and sector-specific job search websites
- Research is required (e.g., Google)
- Assessing relevance based on traffic volume (e.g., Similarweb + metered data)
- At least >5,000 visits per month
- Identifying local versions (.es, /es)

ITM surveys triggered by metered data (specific tasks)

1. Questionnaire design
2. Elaborating a complete list of job search websites.
3. Identifying the URLs corresponding to job applications

	Website	Application URL (example)
#1	Linkedin.com/jobs	Linkedin.com/jobs
#2	infojobs.net/	infojobs.net/candidate/application/2332/apply
#3	ticjob.es/	ticjob.es/esp/ref=232322?status=applied
…	…	

ITM surveys triggered by metered data (specific tasks)

1. Questionnaire design
2. Elaborating a complete list of job search websites.
3. Identifying the URLs corresponding to job applications
4. Discarding unidentifiable events

	Website	Application URL (example)
#1	Linkedin.com/jobs	~~Linkedin.com/jobs~~
#2	infojobs.net/	infojobs.net/candidate/application/2332/apply
#3	ticjob.es/	ticjob.es/esp/ref=232322?status=applied
…	…	

ITM surveys triggered by metered data (specific tasks)

1. Questionnaire design
2. Elaborating a complete list of job search websites.
3. Identifying the URLs corresponding to job applications
4. Discarding unidentifiable events
5. Transforming such URLs into “regular expressions”.

	Website	Application URL (example)	Regular expression
#2	infojobs.net	infojobs.net/candidate/application/2332/apply	infojobs\.net\/candidate\/application\/apply
#3	ticjob.es	ticjob.es/esp/ref=232322?status=applied	ticjob.es\/esp\/\S*?status=applied
…	…		

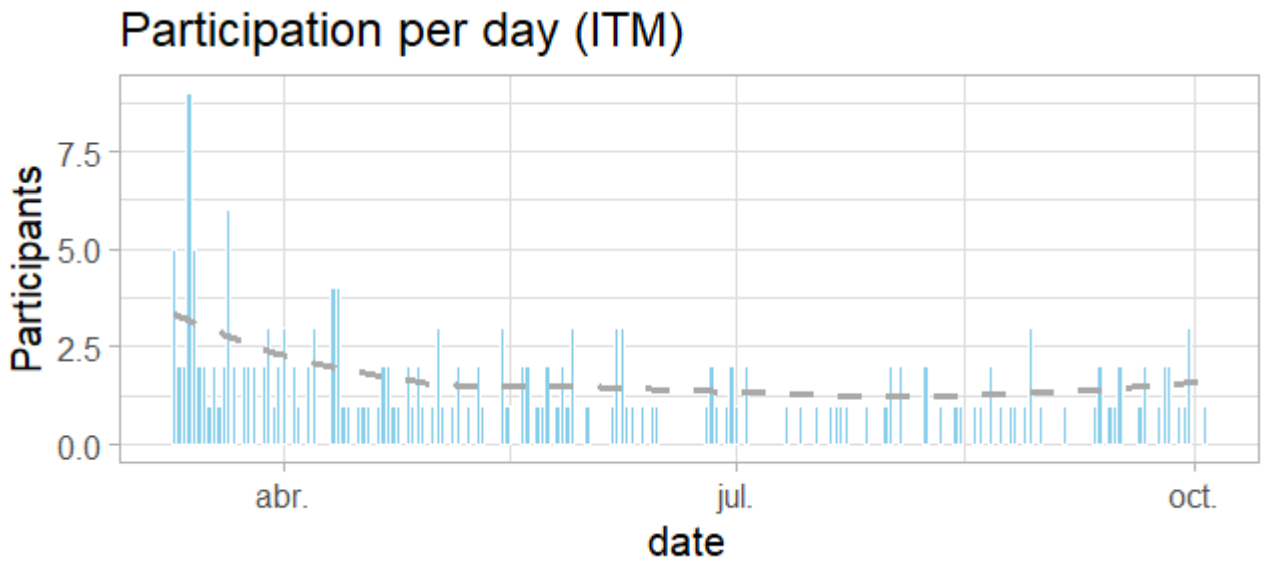
ITM surveys triggered by metered data (specific tasks)

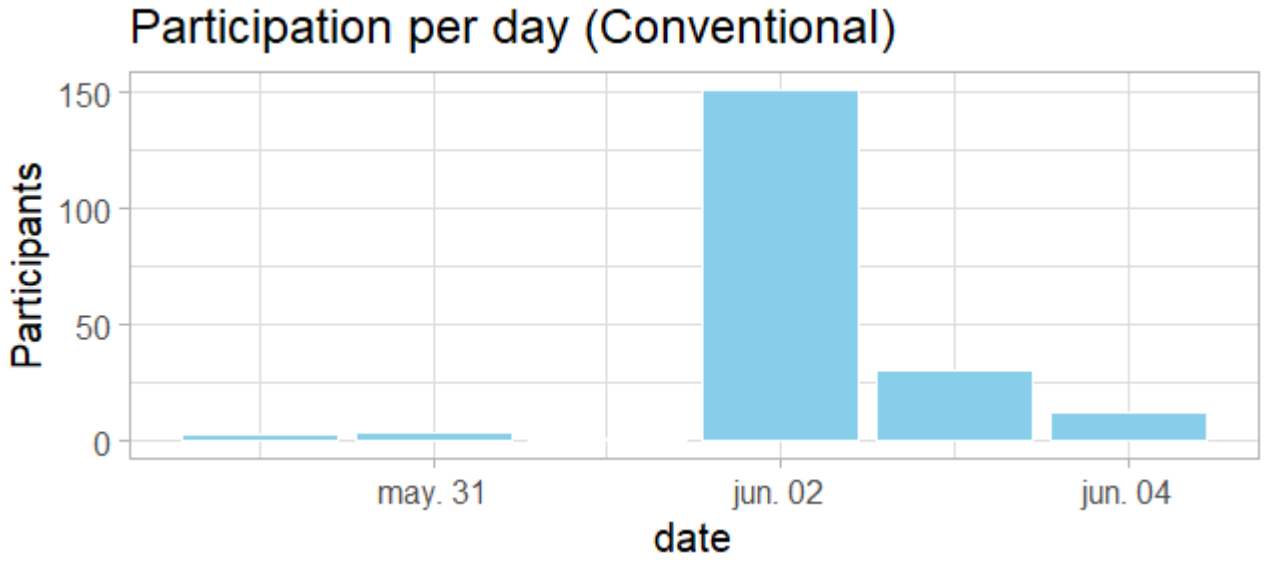
1. Questionnaire design
2. Elaborating a complete list of job search websites.
3. Identifying the URLs corresponding to job applications
4. Discarding unidentifiable events
5. Transforming such URLs into “regular expressions”.
6. Execution through specific software (Web Data Now)

ITM surveys triggered by metered data (specific tasks)

1. Questionnaire design
2. Elaborating a complete list of job search websites.
3. Identifying the URLs corresponding to job applications
4. Discarding unidentifiable events
5. Transforming such URLs into “regular expressions”.
6. Execution through specific software (Web Data Now)
7. Reviewing regularly such URLs, since webpages evolve over time.

Comparing the fieldworks





- All metered and non-metered panelists are randomly assigned to each survey (ITM vs. Conventional) to ensure the inclusion of metered panelists in the conventional survey.
- The ITM survey was launched on March 3rd.
- The conventional survey was conducted over just six days in May-June.
- Metered panelists not used in the conventional survey are reassigned to the ITM survey.
- **207 days** to achieve **198 ITM completes**.

Setting up ITM projects

Specific tasks compared to conventional surveys

Geolocation data

ITM surveys triggered by geolocation data

web data opp

Equivalent tasks

1. Questionnaire design (confirming visits)
2. Elaborating a complete list of beaches.
3. Identifying the location of the beaches on a map tool
4. Discarding irrelevant/problematic locations
5. Transforming beaches into geolocation coordinates

ITM surveys triggered by geolocation data

The effort depends on the numbers of locations of interest

- ≈ 3,500 beaches in Spain
- ≈ 2 min. per beach
- ≈ 120 hours

Risks:

- Including highly visited areas that are not beaches (roads, promenades) could lead to false positives.
- Non-detection caused by the need to clearly separate beaches from other areas
- Precision issues (5-10 meters).

ITM surveys triggered by geolocation data

web data opp

Equivalent tasks

1. Questionnaire design (confirming visits)
2. Elaborating a complete list of beaches.
3. Identifying the location of the beaches on a map tool
4. Discarding irrelevant/problematic locations
5. Transforming beaches into geolocation coordinates
6. Execution through specific software (Web Data Now – geoloc version)
7. ~~Reviewing~~

Conclusions

Summary

- The setup process in ITM surveys is **complex and labor-intensive**.
- **Technology imposes limitations** on data collection, which can only be identified during project setup (e.g., failure to detect linkedin.com/jobs).
- However, **technology is rapidly evolving**. As we implement the job-search project, the meter can now capture in-app data and browser "tab" activity.
- Some existing capabilities (such as HTML gathering) could address certain limitations, albeit at the expense of (1) heightened complexity and (2) privacy concerns.
- ITM frequently results in **prolonged fieldwork durations**.

Promising results

ITM survey triggered by metered data
Completed in 2023

- **High willingness** to participate among panelists already sharing metered data.
  Participation rate of 85.3%
- **No concerns** regarding privacy or perceived intrusiveness were raised.
  Levels of easiness and satisfaction higher than the conventional survey, but mostly explained by being metered panelists.
- Some **positive impacts on data quality**:
  Increased length of answers to open-ended questions: +11.4% to +47.5%.
  Weak effect on explicit non-recall, probably caused by overconfidence in own memory.
- Significant **differences in substantive results**
  Example: estimated prob. of being hired (ITM: 39.6% vs. Conv: 48.3%)
  The time elapsed since the occurrence of the event impacts the substantive answers.

But still open questions

- Can AI provide valuable assistance in specific tasks, like identifying URLs and geolocation coordinates?
- What configurations should we apply to the parameters of the ITM survey? This includes the invitation delay post-event detection and the maximum participation time, to efficiently tackle memory concerns.
- Should these parameters be tailored to the specific research problem?
- What strategies can we deploy to minimize false positives and false negatives?
- How do we assess self-selection bias, especially concerning individuals' inclination to share passive data?

Thanks!

CARLOS OCHOA | RECSM - UPF

Carlos.ochoa@upf.edu

https://www.upf.edu/web/webdataopp

Universitat Pompeu Fabra Barcelona

RECSM Research and Expertise Centre for Survey Methodology

webdataopp

References

Bosch, O. J., Revilla, M. (2022). The challenges of using digital trace data to measure online behaviors: lessons from a study combining surveys and metered data to investigate affective polarization. In SAGE Research Methods Cases. https://dx.doi.org/10.4135/9781529603644.

Ochoa, C. (2022). Willingness to participate in geolocation-based research. PLoS ONE 17(12): e0278416. https://doi.org/10.1371/journal.pone.0278416.

Ochoa, C. and Revilla, M. (2022). Willingness to participate in in-the-moment surveys triggered by online behaviors. Behavior Research Methods (2022). https://doi.org/10.3758/s13428-022-01872-x.

Ochoa, C. and Revilla, M. (2022b). Acceptance and coverage of fast invitation methods to in-the-moment surveys. International Journal of Market Research. https://doi.org/10.1177/1470785322108520 4.

Tourangeau, R. (2000). Remembering what happened: Memory errors and survey reports. In A. A. Stone, J. S. Turkkan, C. A. Bachrach, J. B. Jobe, H. S. Kurtzman, & V. S. Cain (Eds.), The science of self-report: Implications for research and practice (pp. 29–47). Lawrence Erlbaum Associates Publishers.: The Science of Self-report. Psychology Press, pp.29-47.