DESIGN CHOICES IN WP2.2 TIME USE

Version: 13-07-20

In the first advisors meeting the focus was on respondent statistics through in-app feedback and time use subthemes. The second meeting will be focused on the following topics:

- User interface/use experience:
    - Open-ended versus closed-ended entry of activities
        - How to deal with extensive lists of activity categories?
        - A closed list versus open data entry versus hybrid?
    - Type and form of plausibility checks
        - What about checks on number of time activity episodes per day?
        - What about checks on diversity of episodes per day?
        - Force a complete diary?
        - How to deal with incentives and incomplete/less plausible responses?
- Geo-locations:
    - Implement as provisional tracks and stops in the diary?
    - Or only to guide participants?
    - Use as auxiliary data in the estimation stage?
    - How to deal with respondents that refuse location measurements?
- Additional smart features
    - Other mobile device sensors (motion, camera) that may be useful?
    - Mobile device use (apps, contacts, social media) as additional passive/active measurements?
    - Link to wearables such as activity trackers?

DESIGN CHOICES IN WP2.2 TIME USE

Version: 13-07-20

In the first advisors meeting the focus was on respondent statistics through in-app feedback and time use subthemes. The second meeting was focused on the following topics:

- User interface/use experience:
    - Open-ended versus closed-ended entry of activities
      Advice: Find a middle ground option when activity lists get very long, i.e. have a closed list but offer respondents the option to type and combine with a search algorithm.
      Comment: This halfopen option demands for trained machine learning models that match activity descriptions to entries in the lists. Currently, University Utrecht is trying to train such models based on 2016 Dutch HETUS paper diaries that were manually coded. This is not straightforward given that respondents vary in the amount of detail that they provide. Also coders are to some extent subjective and there are coder effects. Nonetheless, this seems a natural option to go when activity lists get very long.
    - Type and form of plausibility checks
      Advice: Be conservative when implementing plausibility checks. Checks relating to missing time slots are acceptable and may even be hard checks. Soft checks may still be acceptable for simple criteria like eating or sleeping, but caution is needed as implausible activity patterns could still be valid. More complex criteria should be avoided as well as criteria based on paradata, such as not allowing respondents to fill in a day with a large time lag. Paradata can be useful, however, for processing of data.
- Geo-locations:
  Advice: Location-time measurements and enriched location data should above all assist respondents. This means that any implementation should be logical to respondents. Active location measurement should, therefore, be tested in a number of user interface options. Passive data collection is easier, but is a black box to respondents and may risk lack of consent. The right balance should be found based on experimentation. Advisors are invited to think of useful designs.
  An influential aspect of time-location measurements are technical issues and battery depletion. The latter implies that the frequency of measurements should be carefully arranged. Also there is a risk of incomparability between respondents.
- Additional smart features
  For this topic little time was left and it will be added to the agenda of future meetings.
  Advice: Passive measurements on smartphone use and wearables' data are definitely very interesting. They should be explored.

SUMMARY WP2.1 MEETING METHODOLOGY

ADVISORS: Florian Keusch
CBS: Marleen Schulte, Barry Schouten
DESTATIS: Karen Blanke
STATBEL: Kelly Sabbe
VUB: Joeri Minnen
UU: Bella Struminskaya

During the meeting two of the selected methodology topics are discussed: feedback of respondent statistics and time use subthemes. The other two selected topics, plausibility checks and geo-locations, are postponed to a future meeting.

1. FEEDBACK OF RESPONDENT STATISTICS

Beforehand the following questions were posed:

- How can we engage respondents for a longer time period?
- How to inform respondents about feedback?
- What feedback options are there?

In general, feedback is an extra feature that may be offered in addition to incentives. Within the ESS, incentive strategies and policies vary greatly across countries. It was decided early on that some design features will be left to the NSI's themselves. The incentive strategy is one such feature. Obviously, the effectiveness of feedback interacts with the incentive strategy, e.g. there may be less to gain when incentives are high. It is anticipated that the feedback strategy should still be detectable as it is randomly varied across countries.

Feedback may serve three purposes:
1) Increase recruitment rates
2) reduce attrition/drop-out rates
3) improve diary data quality
Four groups may be distinguished:
1) those who would otherwise not register
2) those who would otherwise drop out
3) those who would otherwise provide low data quality
4) all others
The question is raised how strong the impact of feedback can be. It is concluded that it must be very salient. Literature is thin, but does not show much evidence of feedback being very impactful. VUB mentions that some of their studies have included feedback, although not very salient in the app itself.

The type of feedback may be a range of different types of individual time use statistics in combination with some form of benchmarking to other persons' time use. VUB shows an example where respondents could afterwards consult a summary of their time use. This was done in a school teachers

study. In the ESSnet, we consider in-app feedback so that the number and range of figures should be limited. Two types of figures may be most interesting: a pie chart of time use on a certain aggregation level, say HETUS first digit, and a bar chart showing time use across the reporting period.
The benchmarking option seems very powerful and leads to two follow-up questions:
1) On what characteristics should we benchmark?
2) Do we use statistics from the on-going study or from a past study?
Tentative conclusions are that the ESSnet field test is too small to go very detailed and to use statistics from the current study. Benchmarking on say age categories and gender seems possible. Decisions are left to the WP2.2 project team.

The options to make feedback very salient were not discussed in detail, but given that invitations go through invitation letters, it is obvious to mention feedback in the invitation letter and to add a leaflet with examples of statistics. At NL it has been proposed to print the leaflet in the shape of a smartphone with the front side a smartphone displaying the statistics example.

The obvious risk of feedback is unnatural respondent behaviour, i.e. respondents are affected. In order to be able to separate the impact from "true" behaviour, feedback may be given instantly or at the end of the reporting period. Furthermore, it can be varied whether it is made salient in the invitation or not.

In the field test, budget is foreseen for a randomization into three feedback groups. In total there are nine groups based on yes/no mention of feedback un the invitation, instant feedback or delayed feedback at the end, and yes/no benchmarking. The groups are:

1. No mention of feedback, no in-app feedback (the control condition)
2. No mention of feedback, instant feedback but no benchmarking at the end
3. No mention of feedback, delayed feedback but no benchmarking
4. No mention of feedback, instant feedback and benchmarking at the end
5. No mention of feedback, delayed feedback and benchmarking
6. Mention of feedback, instant feedback but no benchmarking at the end
7. Mention of feedback, delayed feedback but no benchmarking
8. Mention of feedback, instant feedback and benchmarking at the end
9. Mention of feedback, delayed feedback and benchmarking

Even if budget would suffice to test more than three options, from operational and capacity point of view, we would need to restrict. Apart from the control condition two other options need to be chosen. Since we are interested in the impact on recruitment, the two other options should vary the mentioning of feedback. During the meeting no decision was made.
Proposal: If benchmarking is operationally feasible, we include it in all feedback options, i.e. it is not varied. That leaves options 3, 5, 7, 9 to choose from (or 2, 4, 6, 8 if not feasible). We chose options 7 and 9 (or 6 and 8 when benchmarking is not possible) as we like to investigate the impact on respondents behaviours.

2. SUBTHEMES

Beforehand the following questions were posed:

- General time use (HETUS)
    - Can HETUS be used to derive statistics on mobility, working and media use?
- Mobility - Working hours - Media use
    - How to implement the subthemes?

- How to deal with plausibility checks?
- Do we expect measurement effects (over/underreporting)?
- If so, how to avoid the effects?
- Are large measurement effects expected when moving from a one-time survey to a diary?

This topic was not discussed in full detail, in part because draft subtheme data models are still missing. Per subtheme, it is briefly explained what each subtheme may encompass:

- Transportation: The main activities are tracks and stops. For each track the following information may be asked or measured: what was the purpose, what was the distance, what were the modes of transportation, were there fellow travellers and if so who, did you enjoy the travel, was it a routine travel?
- Working hours: The main activities are limited to work-related activities, including commuting. Non-work-related activities are pooled into a remainder category. The work-related activities may be further detailed and categorized. Additional information may be work location, with or without colleagues, yes/no abroad, work stress, work satisfaction.
- Media use: The main activities are limited to those that involve media. The ICT survey that is conducted in the ESS context may be taken as an example. Additional information may be the purpose of the media use, specifics of the type of device/media, was the media used with others and if so who, activity enjoyment.

During the meeting it was discussed whether there is incentive to go more "smart", i.e. to include measurements instead of, or in addition to, diary questions. For transportation and work, location measurements may be used while for media use the mobile device use itself may be measured. There is some experience, for instance in an IAB/Mannheim study in Germany about this. In general, literature is still thin. Given that the ESSnet itself does not develop tools but tests them (with small modifications/tailoring), most of the options seem better suited for extended phase 2 tests and follow-up HBS –TUS innovation grants. VUB is working on geo-locations and in 2021 they may be included in tests. It is too early to make any statements about this, but it will kept in mind as an option.

In the meeting it is expressed that there is great interest in the impact of diary/longitudinal surveys on the measurement of subthemes that have been surveyed mostly retrospectively.

During the meeting it is proposed that for the next meeting draft data models/questions will be prepared for each of the subthemes. BE will focus more on transportation and work and NL on media use.

Proposal: Prepare subtheme diary studies in detail for next meeting.