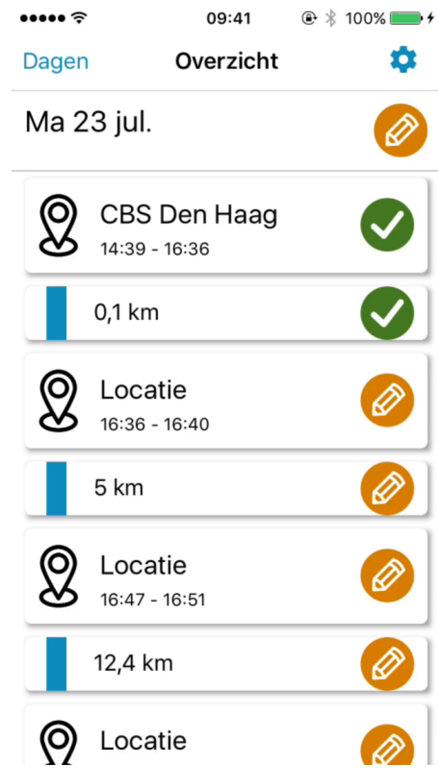
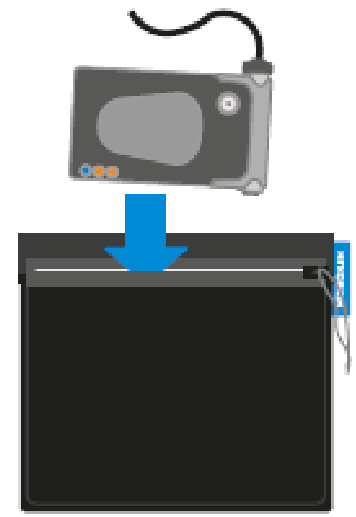




Smart surveys. Case studies in official statistics

Ole Mussmann en Barry Schouten

Seminar Statistiek Vlaanderen, 9 november, 2021

OUTLINE

1. Intro on smart surveys and case studies
2. Demo of case study apps
3. Methodological challenges and early conclusions
4. Discussion

What are smart surveys?

Smart surveys have at least on of the following smart features:

1. Device intelligence: It can use the intelligence (computing, storage) of the device
2. Internal sensors: It can employ the sensors available in the device;
3. External sensors: It can communicate with other sensor systems;
4. Public online data: It can extract publicly available online data;
5. Personal online data: It can go online and request access to existing external personal data;
6. Linkage consent: It can ask consent to link external personal data already in possession of the survey institute.

Smart surveys?

Smart surveys still have the respondent as central point of data collection.

Smart surveys form a bridge between primary (survey) data collection and secondary (big) data collection.

Why smart surveys?

Criteria:

1. Burden: The survey topic(s) are burdensome for a respondent (time or cognitive effort);
2. Centrality: The survey topic(s) are non-central to respondents;
3. Non-survey type: The survey topic(s) do not lend themselves to a survey question-answer approach to begin with;

In other words, to increase response rates and improve data quality.

Third criterion points to reconsideration of concepts and proxy measures

Case studies

- Household Budget Survey: Diary is very burdensome
- Time use survey: Diary is burdensome + recall error
- Travel Survey: Detailed knowledge may be absent
- Physical activity: Burdensome + proxy concepts may be weak
- Living Conditions Survey: Proxy concepts may be weak
- Working conditions: Proxy concepts may be weak

First three projects have dedicated cross-platform apps
Last three projects use external sensor systems

Smart survey case studies

HOUSEHOLD BUDGET SURVEY

Potential features:

- Device intelligence
- Internal sensors
- External sensors
- Public online data
- Personal online data
- Linkage consent

OCR and classification receipts
Camera, location
-
Open streetmaps
Bank transactions data
Scanner data, e-receipts

Smart survey case studies

TRAVEL
SURVEY

Potential features:

- Device intelligence
- Internal sensors
- External sensors
- Public online data
- Personal online data
- Linkage consent

Stopdetetection, ML models
Location, motion, beacons
Activity trackers
Open streetmaps
-
Public transport data

Smart survey case studies

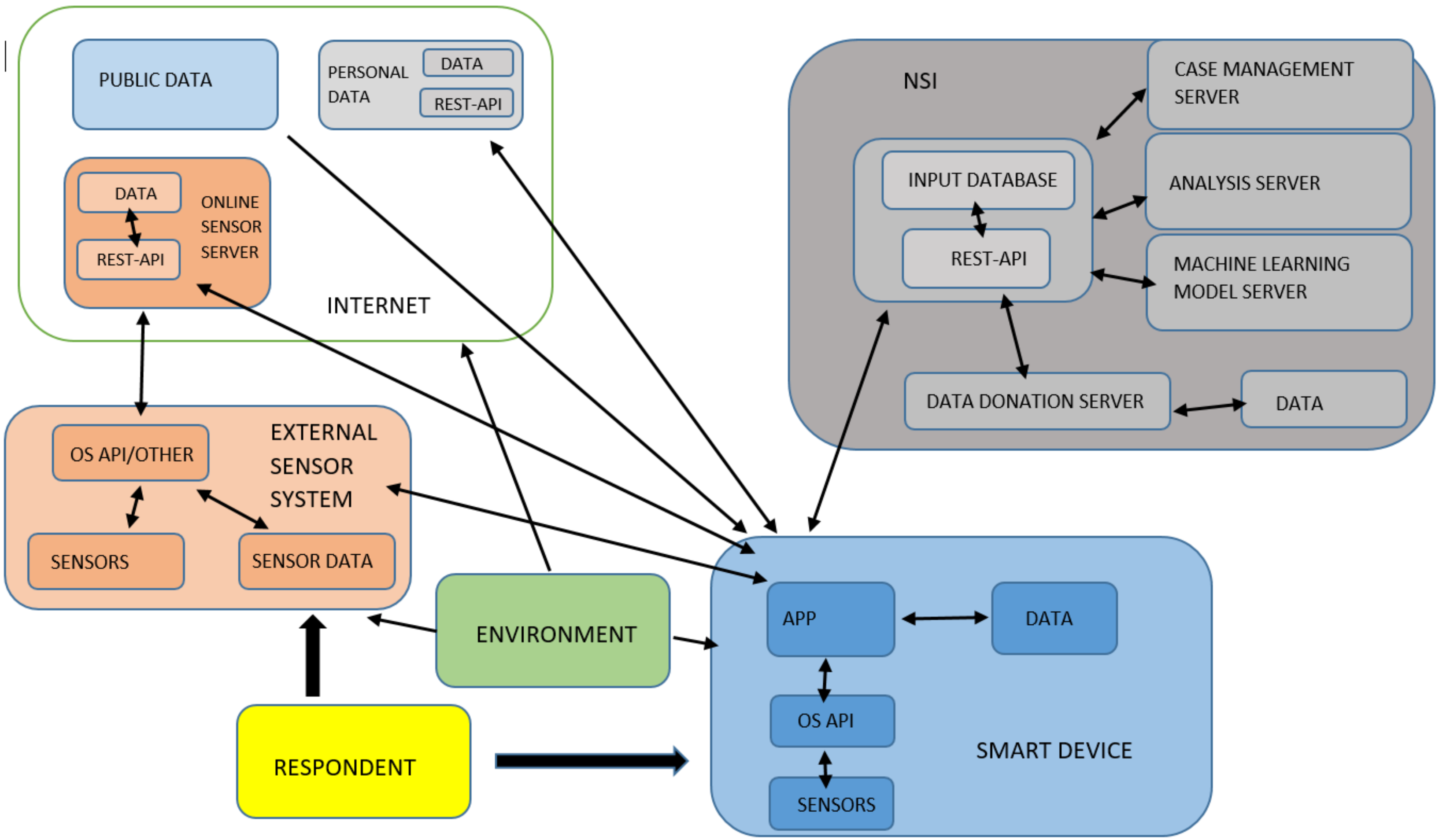
LIVING CONDITIONS SURVEY

Potential features:

- Device intelligence
- Internal sensors
- External sensors
- Public online data
- Personal online data
- Linkage consent

Machine learning models
Camera, light, sound
Indoor climate
Outdoor air quality, streetmaps
Smart meter data
Energy bills

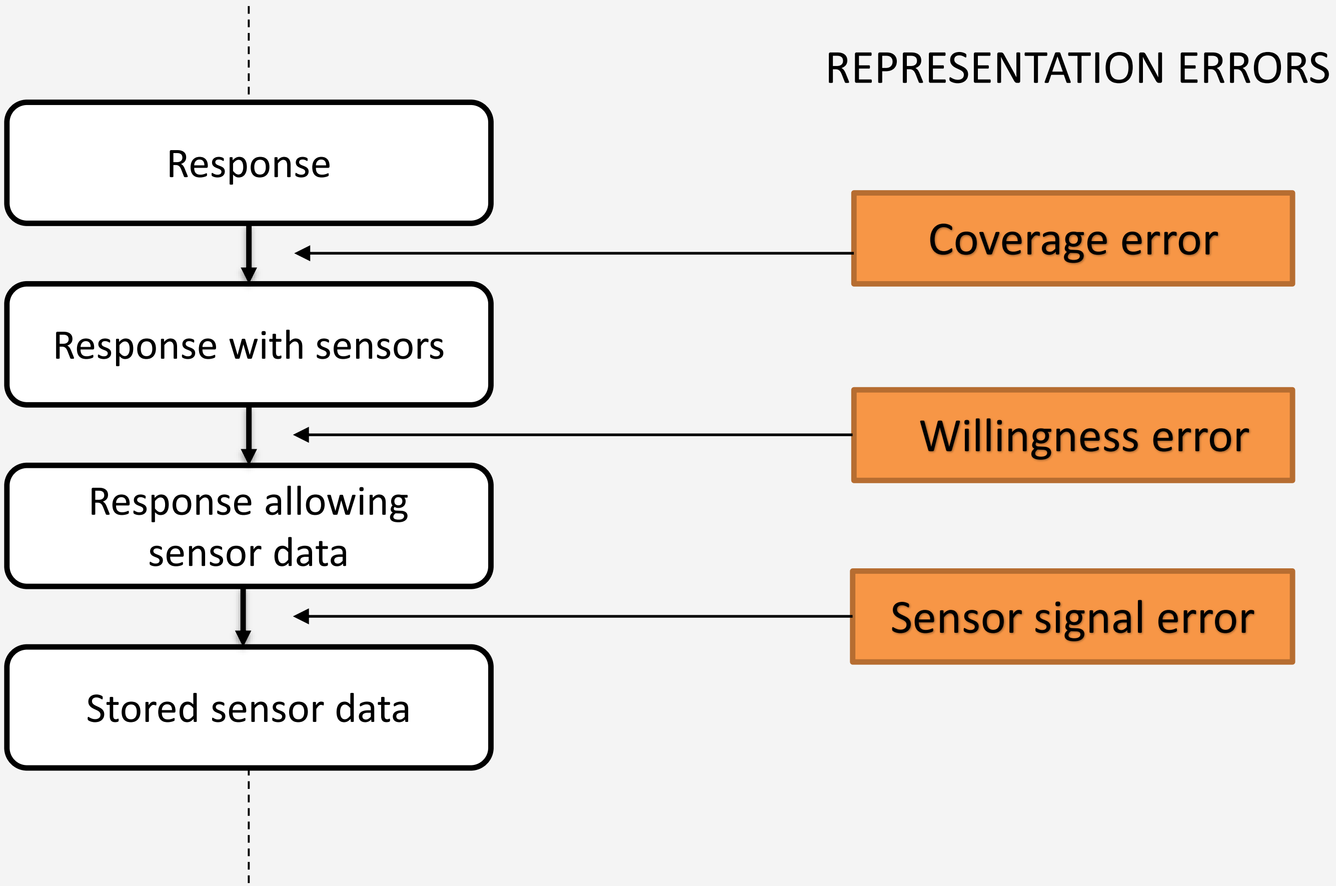
Architecture smart surveys



Measuring and sensing instead of asking?

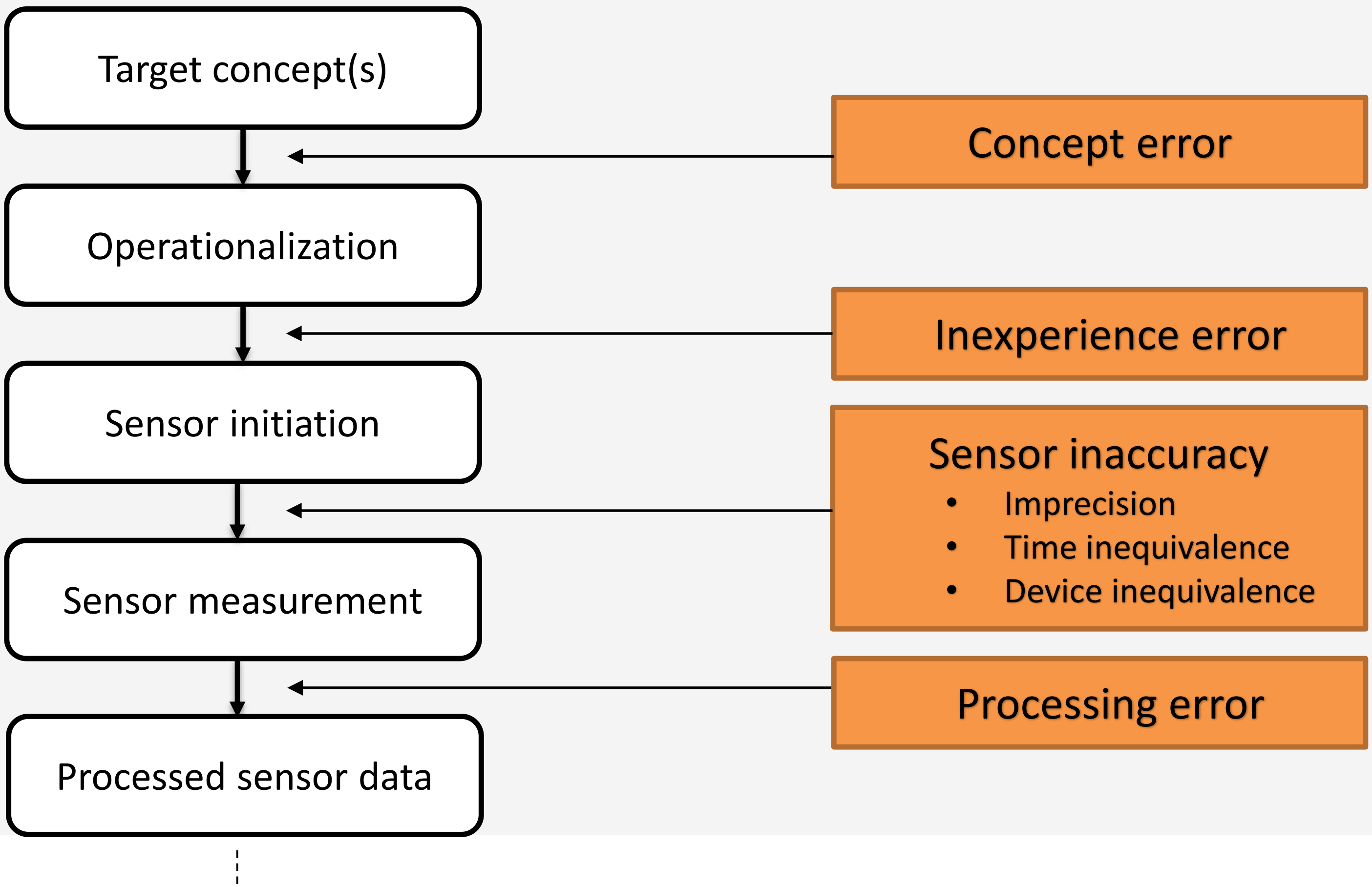
- Sensor and other types of measurements are subject to error as well:
  - Representation
  - Measurement
- GDPR becomes more prominent as respondents do not (always) know the data that is being measured and processed
- Respondent engagement translates to design choices in active versus passive data collection
- Machine learning methods play a dominant role in processing of data and lead to in-device versus in-house trade-offs

Total sensor survey error framework



Respondent engagement and sensor errors - 2

MEASUREMENT ERRORS



Focal points methodology

- User interface/experience: How to integrate different types of measurements/data in one instrument? How to present active-passive involvement?
- Effective data collection strategies and role of interviewers: How to recruit and motivate participants?
- Quality control of alternative sources of data: How to make active-passive trade-offs? Can machine learning predictions be input?

Focal points methodology - continued

- Planned missing designs: Sensor measurements may be collected for a subset of respondents based on heterogeneity and predictability
- Adaptive survey designs: Differentiation may be needed to account for preferences and costs in measurements and/or linkage
- Re-interview designs: Repeated measurements needed to unravel confounding of representation and measurement

Active-passive sensor data collection

Active data collection = Respondents are involved in interpretation of the sensor task, retrieving information through the sensor task, judging the sensor data, and/or submitting the sensor data.

Why active data collection?

1. Respondent engagement: To increase respondent control, to make the survey more enjoyable, to feedback insights;
2. Sensor error adjustment: To adjust for errors that may occur in collecting sensor data;
3. Legal (ethical): To conform to data minimisation principles in data collection legislation (such as GDPR);

DEMO

1. Household Budget Survey (Eurostat):
   - Linked to HBS regulation for ESTAT
   - Period: Two weeks all purchases
   - Specific features: receipt scanning, product search
   - In field test: ES, FI, HU, LU, NL and SI
2. CBS tijdgebruik
   - Linked to HETUS regulation for ESTAT
   - Period: One week all activities at 10 min resolution
   - Specific features: Activity search, future also location tracking and use of NFC tags
3. CBS Onderweg in Nederland
   - Linked to Dutch Travel survey (ODiN)
   - Period: One week all travels
   - Specific features: Location tracking, POI data

DEMO

Household Budget Survey – SV seminarie participants

purchases	photos
17	0
2	0
1	0
13	0
1	0
18	1
1	0
4	1
6	3
42	4
15	4
10	0

Findings/conclusions of smart survey pilots

- User interface/experience
- Effective data collection strategies and role of interviewers
- Quality control of alternative sources of data
- Planned missing designs
- Adaptive survey design
- Re-interview designs

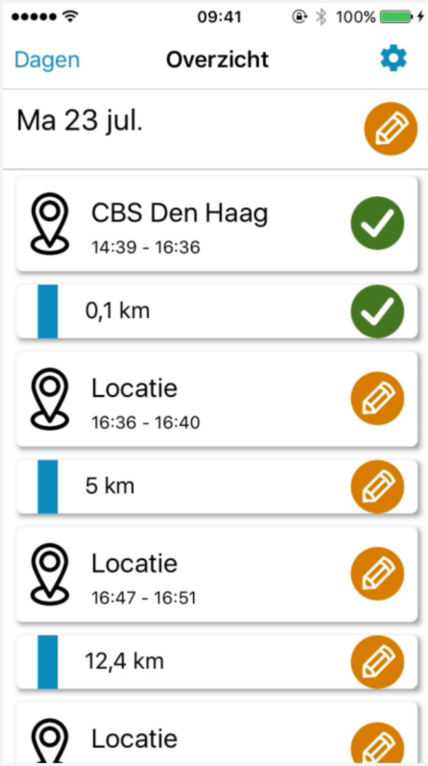
User interface – user experience

- A clear and recognizable UI is key to improve data quality and avoid drop-out. Use well-known look and feel and design options
- UI features:
  - Landing page and instructions
  - Help-options
  - Tutorials
  - Support such as reminder option, helpdesk contacts
- Clear presentation of what respondents need to check and validate
- Personalized insights

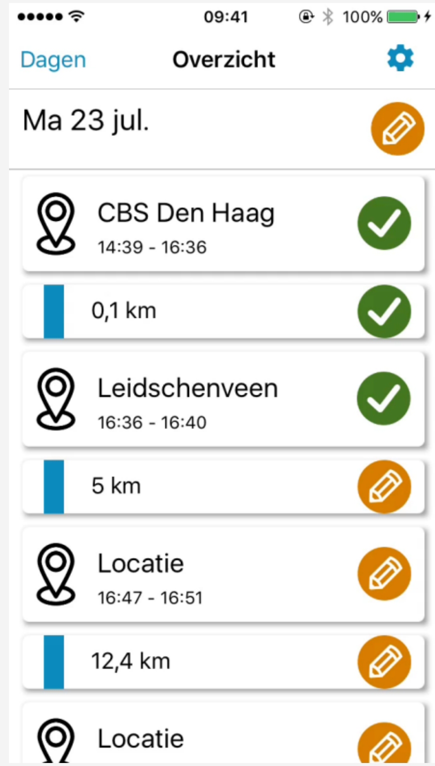
Usability can be analyzed using paradata on in-app navigation

User interface - Travel survey case study

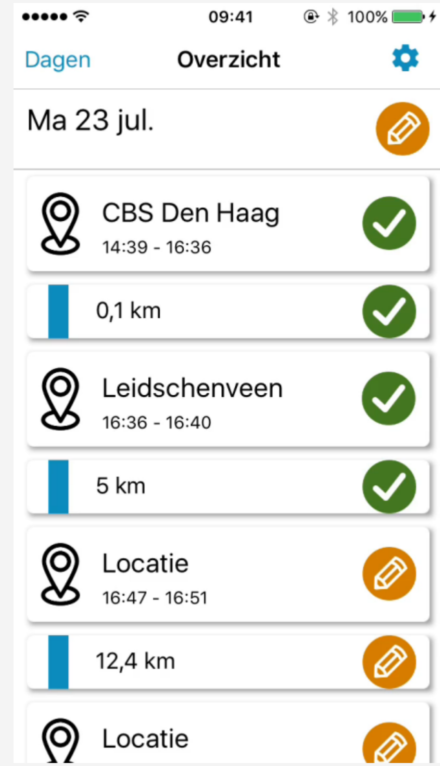
Trip purpose



Trip transportation mode



Daily questions



User interface – in-app paradata

userName	action	objectName	time
1234567	open screen	LoginScreen	2021-09-10 13:29:25
1234567	open screen	Menu	2021-09-10 13:30:11
1234567	open screen	StartQuestionnairePage	2021-09-10 13:30:11
1234567	close screen	StartQuestionnairePage	2021-09-10 13:34:59
1234567	openTab	ReceiptListScreen	2021-09-10 13:35:50
1234567	openTab	ReceiptListScreen	2021-09-10 13:35:56
1234567	openTab	OverviewScreen	2021-09-10 13:35:58
1234567	openTab	ReceiptListScreen	2021-09-10 13:36:00
1234567	open screen	ManualEntryScreen	2021-09-10 13:36:20
1234567	open screen	SearchWidget	2021-09-10 13:36:23
1234567	close screen	SearchWidget	2021-09-10 13:36:29
1234567	open screen	DiscountDialog	2021-09-10 13:37:14
1234567	close screen	DiscountDialog	2021-09-10 13:37:22
1234567	open screen	DatePicker	2021-09-10 13:37:44
1234567	close screen	DatePicker	2021-09-10 13:38:49
1234567	open screen	ManualEntryInformationIncompleteDialog	2021-09-10 13:39:04
1234567	close screen	ManualEntryInformationIncompleteDialog	2021-09-10 13:39:11
1234567	open screen	SearchWidget	2021-09-10 13:39:19
1234567	close screen	SearchWidget	2021-09-10 13:39:23
1234567	open screen	SearchWidget	2021-09-10 13:39:41
1234567	close screen	SearchWidget	2021-09-10 13:39:45
1234567	showTutorialManualEntry	InkWell	2021-09-10 13:40:15
1234567	nextTutorialPage	InkWell	2021-09-10 13:40:33
1234567	nextTutorialPage	InkWell	2021-09-10 13:40:38
1234567	open screen	ClosingScreenWarning	2021-09-10 13:40:42

Search for time use activities – HBS products

Formal classifications (HETUS - COICOP):

- official language
- not always intuitive to respondents
- many ‘other’ categories

Product lists: common language, including common typos, and inclusion of brand names
Activity lists: separate main – side activities, common language,

Search algorithms match based on number of edit steps to products in the list, possibly weighted for position in the word

Recruitment and motivation strategies

- Use of incentives: Current strategy is to give a small unconditional incentive and a larger conditional incentive
- Interviewer support: Interviewers are powerful in recruiting respondents and also act as helpdesk. They need to know the app and the basic issues.
- Landing page and instructions: Landing page is rarely visited
- Helpdesk: Important to have an in-app option
- Personalized insights: No strong evidence that it helps increase response rates, but respondents do expect it

Complication: What is a ‘complete’ response?

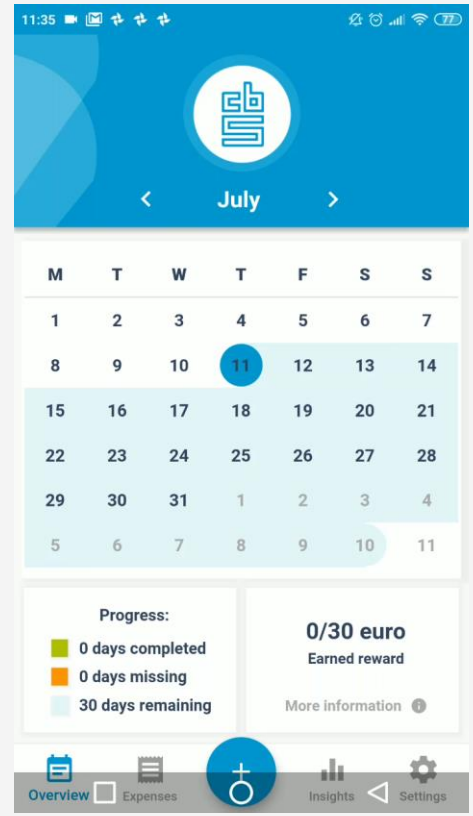
Recruitment strategies - Household Budget Survey

How to increase participation?

Provide individual feedback on household expeditures:

- Without a benchmark
- With a benchmark

In order to reduce impact of study on household behavior, the feedback can be delayed for a specified time period



Recruitment strategies – Travel Survey response rates

Gender	Registered	>7 days diary
Female	34%	25%
Male	33%	22%

Etnicity	Registered	>7 days
Native	36%	26%
Non-western	22%	12%
Western	27%	16%

Age	Registered	>7 days
16-25	42%	28%
26-45	39%	31%
46-65	33%	23%
65+	20%	13%

Urban	Registered	>7 days
Not	32%	23%
Little	35%	23%
Moderate	34%	24%
Strong	33%	26%
Very strong	33%	23%

Recruitment strategies – Living conditions survey (Ilic et al 2021)

	Favourite place		Heating device	
	n	(%)	n	(%)
Unwilling				
Privacy concerns	202	56.4	58	15.1
Don't want or feel like it/unnecessary	26	7.5	14	3.7
Unable or incapable				
Technical problems with survey	40	11.2	39	10.2
Respondent's related technical issues	13	3.6	8	2.1
Do not have favourite place	20	5.6		
Unreachable*			196	51.0
Not sure how home is heated			1	0.3
Not at home	13	3.6	11	2.9
Other reasons	17	4.8	15	3.9
Did not provide a reason	26	7.3	42	10.9
Total	358	100	384	100

Active-passive data collection and quality control

- Sensor data and other types of measurements are imperfect, i.e. communication with respondents often necessary
- Sensor data and other types of measurements insufficiently rich and additional contextual data is imperative
- To obtain trust and create transparency, active engagement of respondents are often paramount regardless of quality

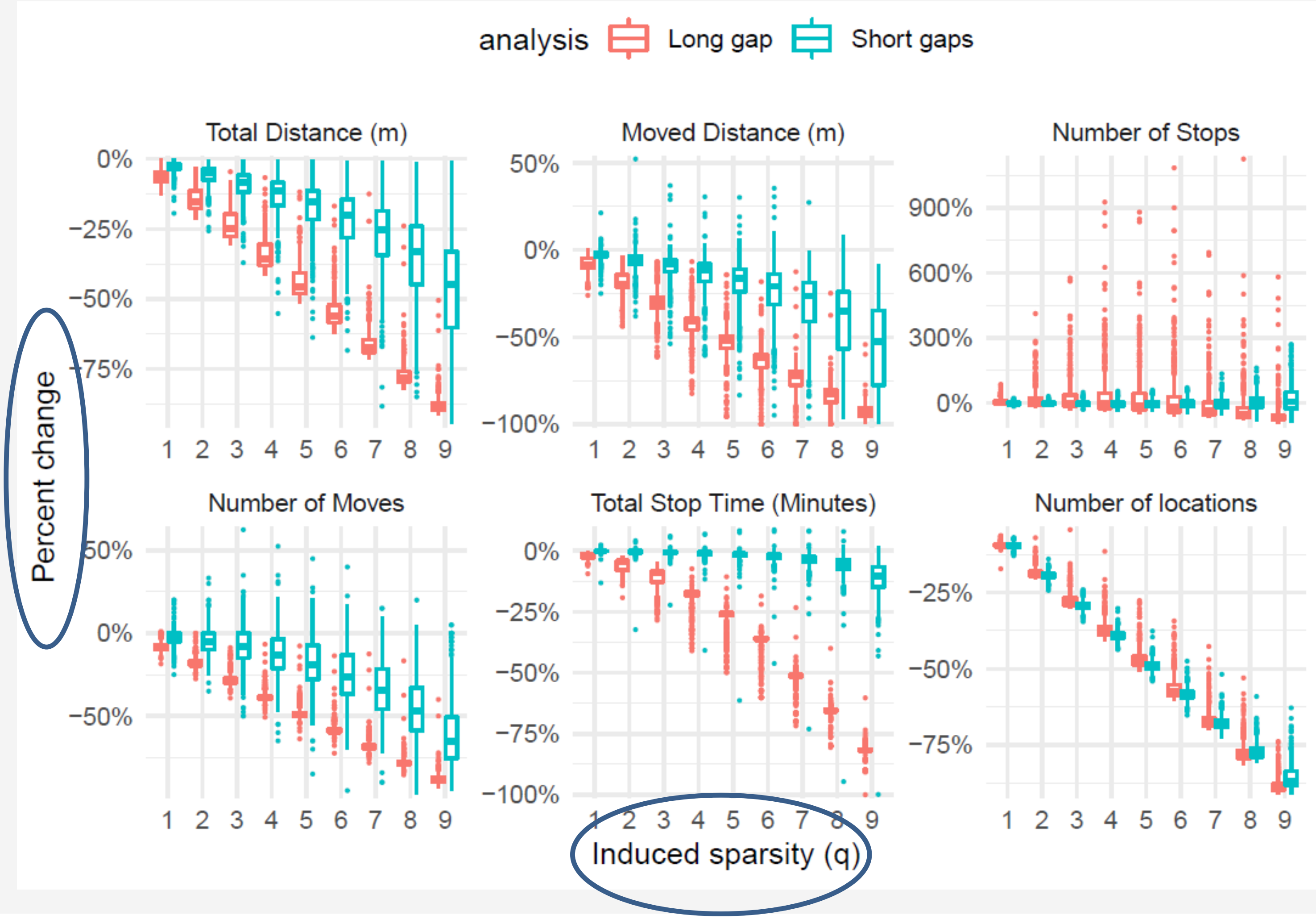
Leads to question to what extents data processing must be done in-device and in-house

Active-passive trade-off - transport mode prediction (Smeets 2019)

Features: time-location data, derived data, respondent background

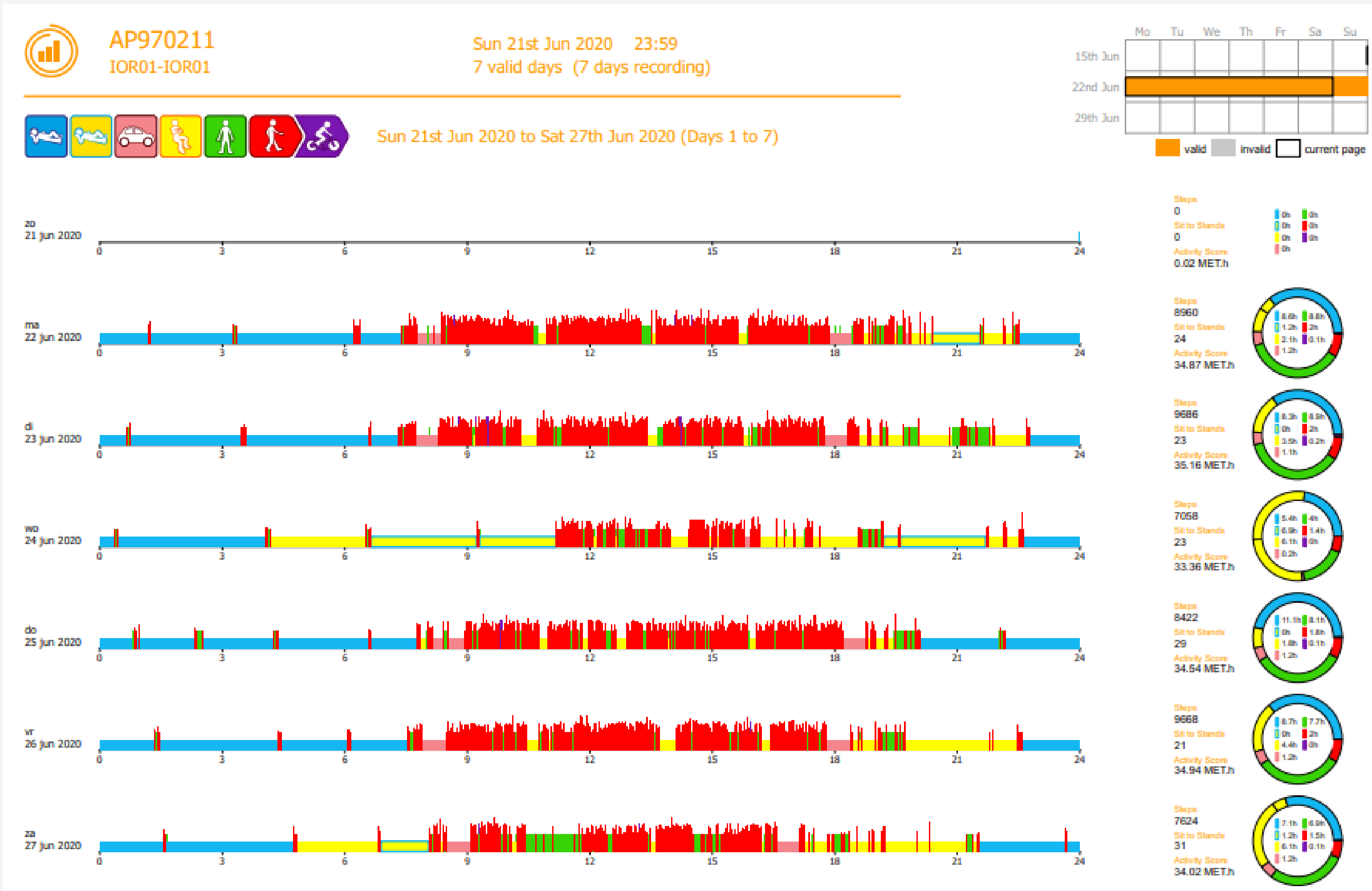
predicted	observed									
	E-bike	bike	car	metro	bus	scooter	train	tram	User error	walk
E-bike	70	164	96	0	2	1	1	0	0	22
bike	29	361	51	0	2	0	0	0	1	35
car	8	20	1308	0	8	3	5	0	3	18
metro	0	11	24	13	0	1	7	2	0	9
bus	4	20	199	0	24	1	4	0	1	5
scooter	13	14	195	0	0	22	0	0	0	4
train	2	4	74	0	2	0	142	0	1	10
tram	2	53	35	1	4	0	7	15	2	35
User error	10	54	109	0	1	1	8	2	16	91
walk	10	59	82	0	2	0	1	1	3	671

Active-passive trade-off - missing location data (McCool 2021)



Stat Vlaanderen, 09-11-21

Active-passive trade-off - Predicted physical activity (Luiten 2021)



HBS – Receipt processing

Steps in receipt processing:

1. In-app scan including info on light and contrast
2. In-app OCR and NLP to interact with respondent
3. In-house OCR and NLP
4. In-house classification
5. (optional) In-house manual check
6. In-app feedback and validation

Options in classification:

- Direct training of ML models using annotated receipts
- Training of models through EAN/GTIN product descriptions
- Match to receipt texts provided by shops

Planned missing designs

Motivation: Sensor measurements may be expensive and may be more/less relevant for different subpopulations.

Second sample stage may be introduced based on:

- Baseline covariates (sampling frame, linked auxiliary data)
- Recruitment survey variables

Subpopulations that are least predictable, i.e. show largest heterogeneity, may be oversampled.
E.g. young, single household persons with average BMI in activity tracking

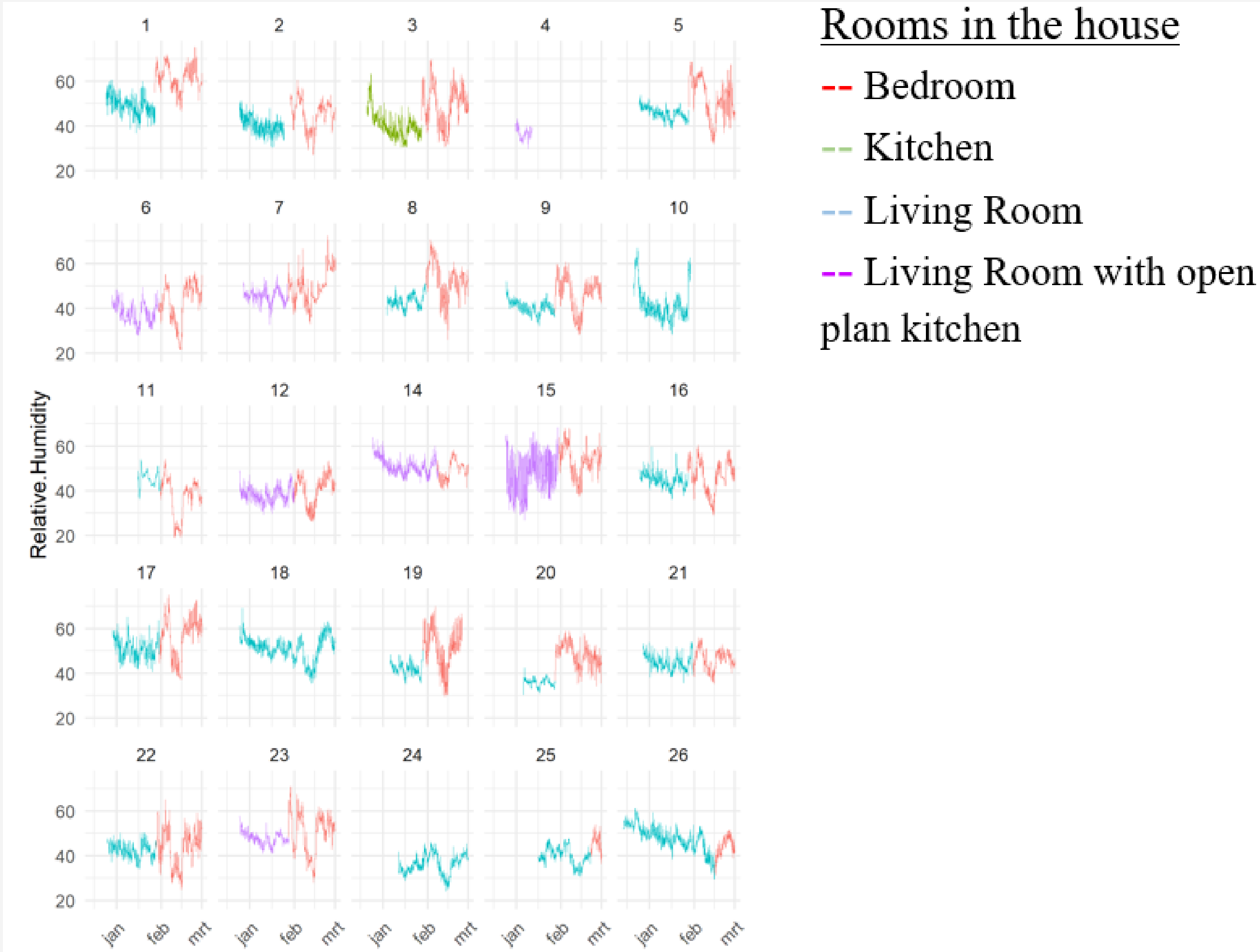
Planned missing design- Living conditions survey

Indoor climate systems:

- App based.
- The data are downloadable in raw format by respondents and/or owner of the climate systems
- CO2, CO, PM2.5, VOCs, NO2, 03, Air pressure, Temperature, Relative humidity

- What dwellings/households should be oversampled?
- How long shall we set measurement periods?
- How to sample locations in a dwelling?

Planned missing designs - Living conditions survey



Adaptive sensor survey designs

Motivation: Persons may be more/less able, more/less willing to enable sensors, and/or more/less prone to break-off.

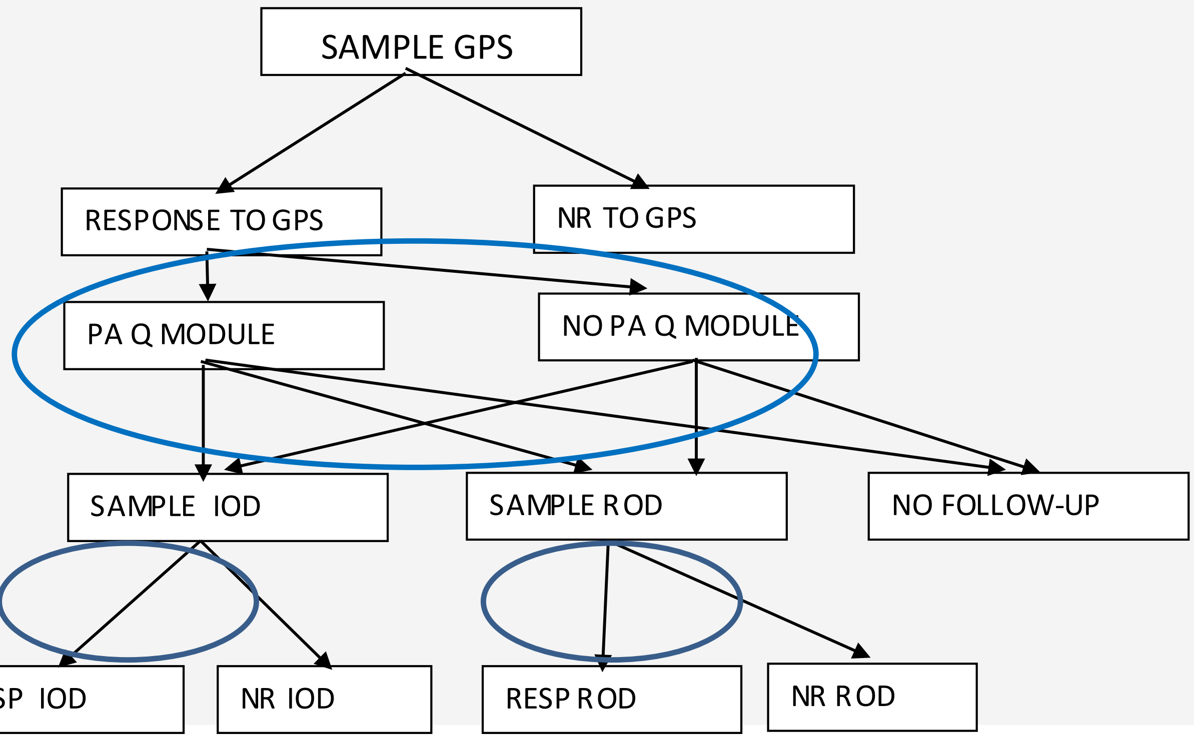
Design features:

- Mixed-devices
- (Push-)Notifications and reminders
- User interface look and feel
- Incentive and feedback to respondent
- Soft and hard edit rules

Interesting respondent features are those that relate to both sensor concepts and competence/willingness/break-off.
E.g. age, BMI, eating/drinking/smoking habits in health

Physical activity pilot – adaptive survey design

Choice of PA measurement and IOD/ROD recruitment effort



Physical activity pilot - quality-cost priorities

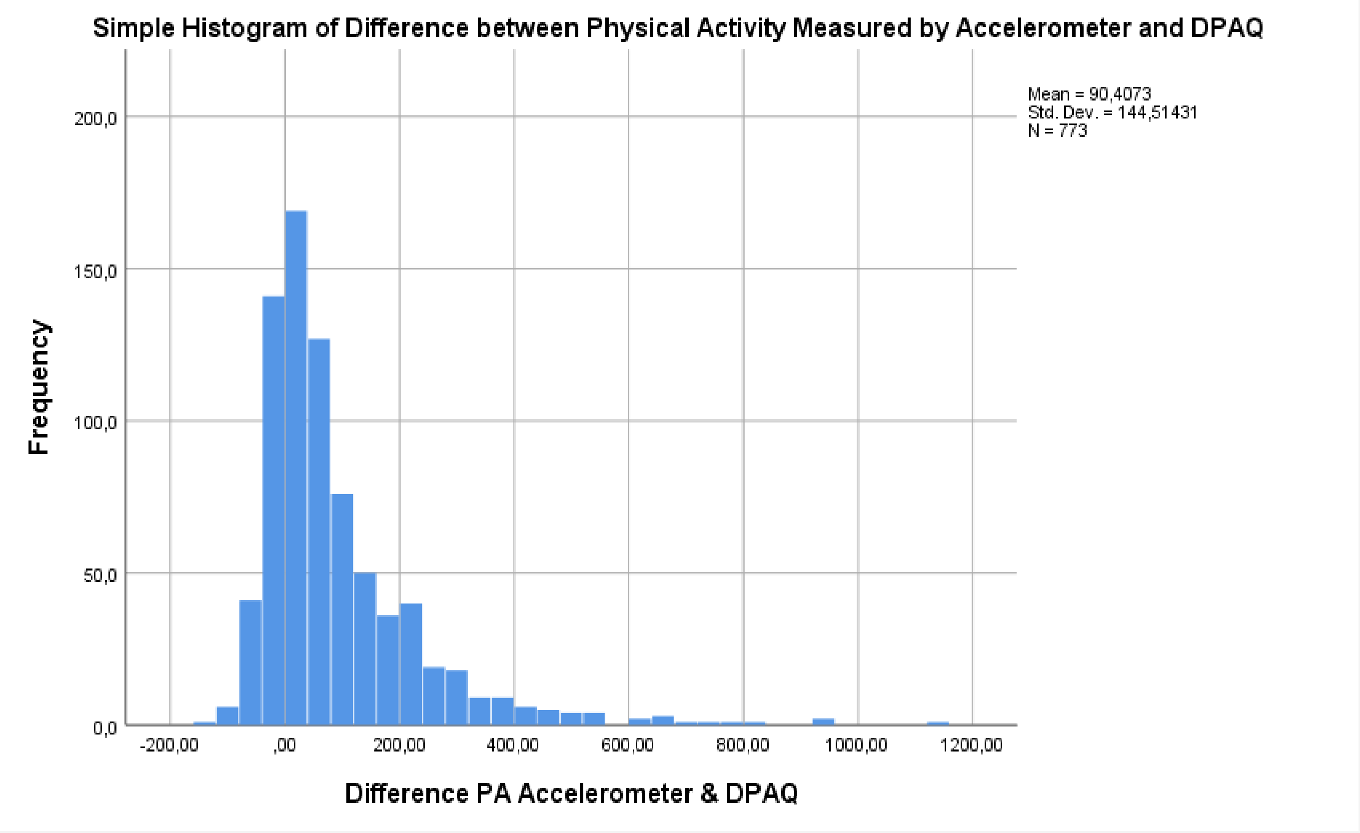
- Objective: Minimize relative measurement bias on PA
- Set lower bound to representation
- Set lower bound to precision of estimated PA
- Set upper bound on maximum difference in relative measurement bias between relevant population subgroups

PA as measured by IOD is considered to be measurement benchmark

Strata from a CHAID classification of PA per week

1. Very good self-reported health and male: 459 minutes per week
2. Very good self-reported health and female: 370 minutes per week
3. Good self-reported health and low BMI: 422 minutes per week
4. Good self-reported health, average to high BMI and male: 383 minutes per week
5. Good self-reported health, average to high BMI and female: 302 minutes per week
6. Good self-reported health and high BMI: 295 minutes per week
7. Fair or bad self-reported health: 255 minutes per week

Relative ME bias PA (moderate - vigorous per day)



Recruitment rates and relative measurement bias

Health	Gender	BMI	RR	ΔME
Very good	Male	All	68%	91 min
Very good	Female	All	74%	88 min
Good	All	Low	79%	77 min
Good	Male	Average	76%	81 min
Good	Female	Average	80%	99 min
Good	All	High	74%	91 min
Fair-bad	All	All	83%	103 min

- Stratum 1, 2 and 6 underrepresented
- Stratum 5 and 7 have largest relative ME bias
- Relative ME bias much larger than variation in relative ME bias

Conclusions

- Smart surveys employ the functions/features of smart devices
- They complicate infrastructure and logistics and are especially interesting when surveys are prone to measurement error and/or high burden
- Smart surveys make complicated trade-offs in active – passive data collection, i.e. between a high and balanced response and data quality
- Smart survey methodology ranges from user interface design to recruitment and motivation strategies to adaptive survey designs to advanced estimation strategies to efficient planned missing designs
- Especially machine learning plays a dominant role in smart surveys

Discussion

- Active-passive data collection design choices
- Choices between in-device and in-house processing using machine learning methods
- Recruitment for sensor system studies from within existing surveys, e.g. EHIS or BHIS
- New roles/tasks of interviewers
- GDPR implications with respect to data minimisation