



REPUBLIC OF SLOVENIA
STATISTICAL OFFICE RS



Universiteit
Utrecht



UNIVERSITÄT
MANNHEIM



Smart Survey Implementation Project results - Methodology

Grant Agreement Number: 101119594 (2023-NL-SSI)

MASS workshop, 4-5 June 2026, Barcelona

*Disclaimer: Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Eurostat.
Neither the European Union nor the granting authority can be held responsible for them.*

SSI consortium

Maren Fritz & Florian Keusch, **University of Mannheim, Germany**



Johannes Volk, Jerome Olsen, Lasse Häufiglückner & Karen Blanke, **DESTATIS, Germany**

Claudia de Vitiis, Fabrizio de Fausti, Francesca Inglese, Marco Terribili, Lino Fazio & Monica Perez, **ISTAT, Italy**

Joeri Minnen, Pieter Beyens, **Hblts, Belgium**

Theun-Pieter van Tienoven, **Vrije Universiteit Brussel, Belgium**



Patrick Lusyne, **Statistics Belgium, Belgium**

Danielle McCool, Peter Lugtig & Bella Struminskaya, **Utrecht University, the Netherlands**

Remco Paulussen, Anne Elevelt, Jelmer de Groot, Maaïke Kompier, Barry Schouten, Jonas Klingwort & Janelle van den Heuvel, **Statistics Netherlands, the Netherlands**

Julie Solard, Simon Quentin, Fanny Mikkol, **INSEE, France**



Universiteit Utrecht

Nina Berg & Aina Holmøy, **SSB, Norway**

Mateja Zgonec, **Statistics Slovenia**



REPUBLIC OF SLOVENIA
STATISTICAL OFFICE RS



Smart diary studies

The screenshot shows the TIMELOG app interface. At the top, the status bar displays 9:41 AM and 100% battery. The app header includes a back arrow and the title 'TIMELOG' with a clock icon. The main form consists of several input fields with green text for recorded data: 'What time did your activity take place?' (19 Feb 2024, 07:14 - 07:23), 'What was your primary activity?' (900. Travelling to/from work), 'What was your secondary activity?' (820. Listening to live-broadcasted radio (also digital)), 'How did you travel?' (By foot), 'Starting point' (Home (My Locations)), and 'End point' (N 16 31). At the bottom, there is a question 'Did you use a device during this activity?' with 'Yes' and 'No' buttons, and an orange 'Add' button.

Time use

The screenshot shows the 'Overzicht' (Overview) screen of a household budget survey app. The status bar at the top shows 11:14. The header displays 'Overzicht' and a notification 'Nog 3 dagen wachten' (Still waiting 3 days) with a clock icon and the text 'Tot de start van het onderzoek' (Until the start of the research). Below this is a calendar for 'januari 2025' with days of the week (ma, di, wo, do, vr, za, zo) and dates (20, 21, 22, 23, 24, 25, 26). The selected date is 'Maandag 20 januari'. The section 'Totaal uitgaven' (Total expenses) shows '2 Producten of diensten' (2 Products or services) and a total of '€13.50'. There is a link 'Uitgaven bekijken' (View expenses) with a right arrow. Below this are two buttons: 'Uitgave toevoegen' (Add expense) in blue and 'Dag voltooien' (Complete day) in white. At the bottom is a navigation bar with icons for 'Overzicht' (home), 'Bonnen' (list), 'Bon Toevoegen' (plus), and 'Settings' (three dots).

Household budget survey

Workpackage 2 – smart data in diaries

9:41 AM 100%

TIMELOG

What time did your activity take place?
19 Feb 2024, 07:14 - 07:23

What was your primary activity?
900. Travelling to/from work

What was your secondary activity?
820. Listening to live-broadcasted radio (also digital)

How did you travel?
By foot

Starting point
Home (My Locations)

End point
N 16 31

Did you use a device during this activity?

Yes

No

Add

Geolocation
Geolocation demonstration

Start date
15/09/2024

Show stops

Show path

Show tracks

Path

Geotracking using location sensors

Time use

11:14

Overzicht

0 3 Nog 3 dagen wachten
Tot de start van het onderzoek

januari 2025

ma di wo do vr za zo

20 21 22 23 24 25 26

Maandag 20 januari

Totaal uitgaven
2 Producten of diensten

€13.50

Uitgaven bekijken

Uitgave toevoegen

Dag voltooien

Overzicht Bonnen Bon Toevoegen Settings

Household budget survey

DE SCHORREN

Pro forma ticket - sales
(tussenrekening)

Lightspeed ticket nummer: 39231

12 15/12/2023 18:20

#1 Linda

#	BTW	Naam	Prijs
1 A		Ice Tea	3.00
1 B		Steak 230gr	26.00
1 B		* Champignon	3.00
1 B		CurryWorsten	10.00
2 A		Ice Tea Zero	6.00
1 A		Val Lime	2.70
Totaal			€ 50.70
12%		Netto	34.82
21%		BTW	4.18
		Totaal	39.00
			2.03
			11.70

De Schorren
Graspolderlaan 32
2660 Hoboken
+32495902643
BE0456974126
www.deschorren.be

Controlegegevens
Datum: 15/12/2023
Tijdstip: 19:15:35
Counter: 475392/634407 PS
Signatuur:
ID: BMC04145446
Kaart ID: 0456974126-001
PLU Hash: 19939a85
Ticket nummer: P124518
Versie: Version 8.0.200692 Lightspeed Server
Productienummer: BPOS0025003243
Device ID: 0DBFD7AC-3A3D-4A3D-8C70-FC5EE058E448

Dit is geen geldig BTW-kasticket.

Receipts using camera

Structure of Smart Survey Implementation

- Workpackage 1: coordination
- Workpackage 2: Methodology
- Workpackage 3: IT/microservices
- Workpackage 4: Business process
- Workpackage 5: legal/ethics

Flow diagram: example for Household Budget

Methodology

IT/ microservices

1. Respondent installs
Smart app

2. Takes pictures of
receipts

3. R checks products
& amounts

4. Further annotates/
answers questions

Doc understanding &
labeling

OCR of text

Match to Coicop
categories



Flow diagram

Methodology

1. Respondent installs Smart app

2. Takes pictures of receipts

3. R checks products & amounts

4. Further annotates/ answers questions

IT/ microservices

Doc understanding & labeling

OCR of text

Match to Coicop categories

WP 2: Methodology

Task 1: How do we recruit respondents successfully for smart surveys?

Task 2: How should we use Machine Learning to use smart data successfully?

Task 3: Issues of User Experience & User interaction. How do we let respondents interact with smart data?

Task 4: How do we combine data from smart surveys with traditional data (web or paper diaries)?

Workpackage 2: Methodology

Deliverables on CROS:

<https://cros.ec.europa.eu/book-page/smart-surveys-wp-2-methodology>

Results today based on

1. Field experiments in Norway, Belgium, Germany, France
2. User-tests in Italy, Germany, France, the Netherlands, Norway, Belgium

Task 1: Recruitment for Smart Surveys

- Success of smart surveys highly depends on participant commitment
- Aim: ensure participation of sample close to population to obtain reliable and valid data for official statistics in Europe
- Goal: refining and improving recruitment strategies for smart surveys
 - Simplifying data collection by utilizing smart features for documenting daily activities or scanning receipts → reducing effort required for participation
 - But mentioning of smart features may raise privacy concerns

SSI project – Large Field Tests

Country	Survey	Field period	Sample size (gross)	Experimental approach
Belgium	Time Use Survey (7- and 2-day diary)	Apr-June & Sep-Dec '24	6,313	Survey invitation, Mode flexibility as follow-up
Germany	Household Budget Survey (14-day diary)	Nov-Dec '24	7,049	Survey invitation
Norway	Household Budget Survey (7-day diary)	Apr-May '24	1,973	CATI recruitment, Download platform

Participation Rates

Country	Gross sample (n)	Completed interviews (n)	Break-offs (n)	App download rate (ADR; %)	Break-off rate (BOR; %)	Response Rate 1 (RR1; %)
Belgium	1,269 (7d)	28 ^a	93	9.5%	76.9%	2.2%
	5,044 (2d)	118 ^a	420	10.7%	78.1%	2.3%
Germany	7,049	126 ^a	152	5.6%	68.3%	1.8%
Norway	1,973	444 ^b	66	25.3%	11.2%	22.5%

^aCases in Belgium and Germany were counted as completes if activities or expenses were entered on all days of the diary.

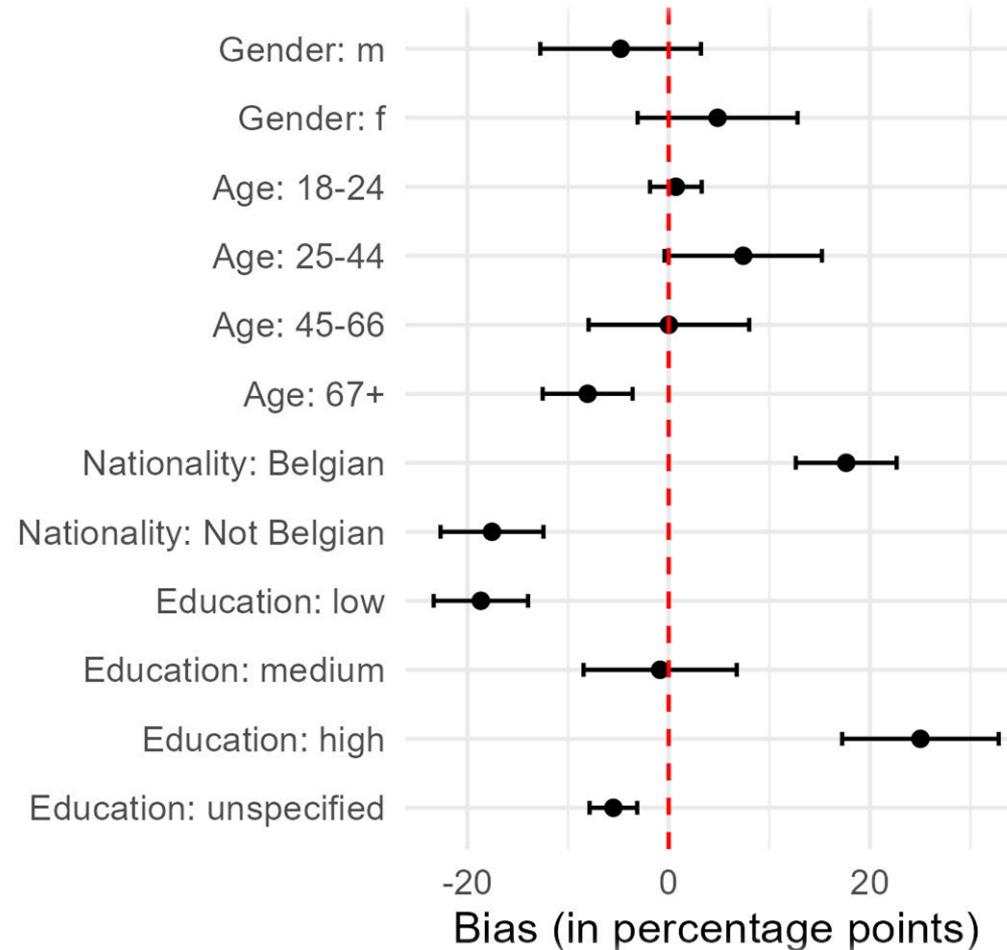
^bCases in Norway were counted as completes if expenses were entered on at least one day of the diary.

$$ADR = \frac{I + P + R_{2,12}}{(I + P) + (R + NC + O) + (UH + UR + UO)} \quad BOR = 1 - \frac{I}{I + P + R_{2,12}} \quad RR1 = \frac{I}{(I + P) + (R + NC + O) + (UH + UR + UO)}$$

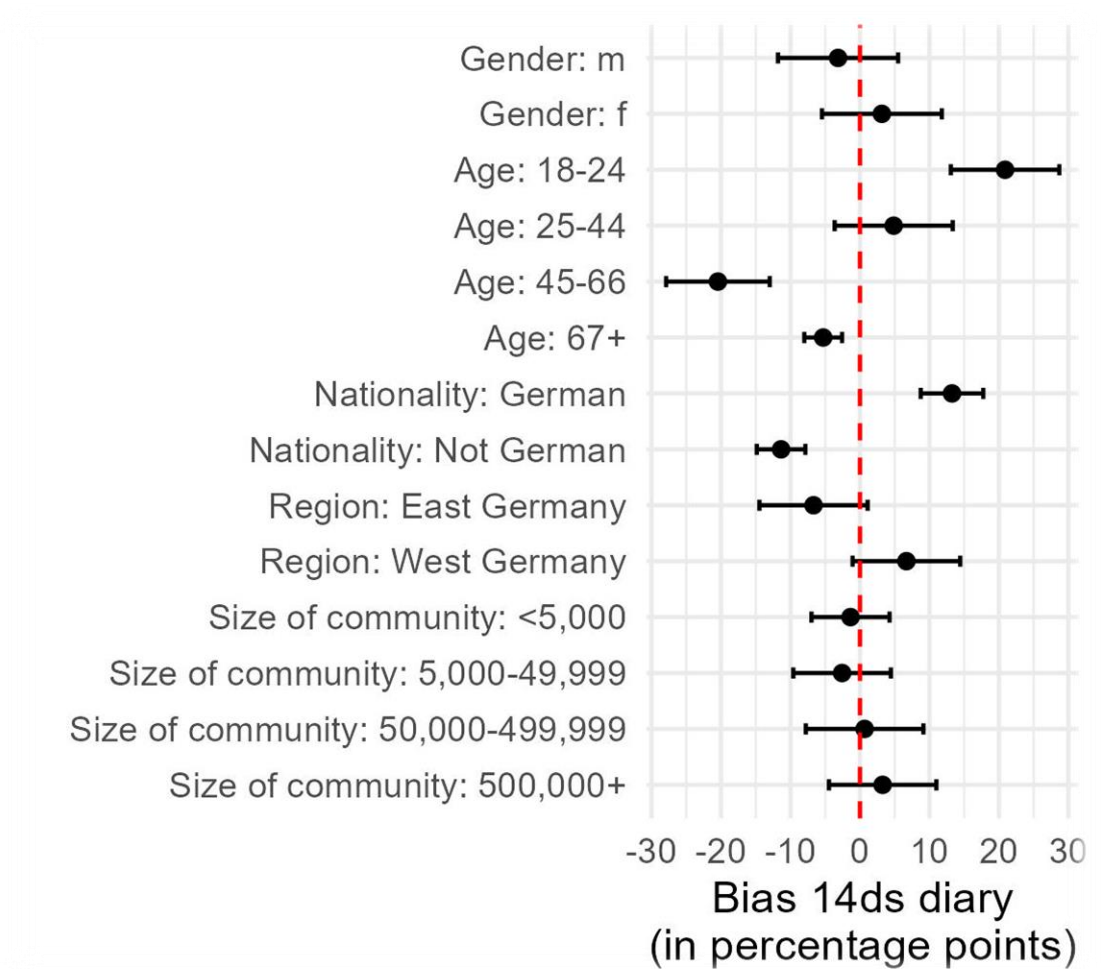
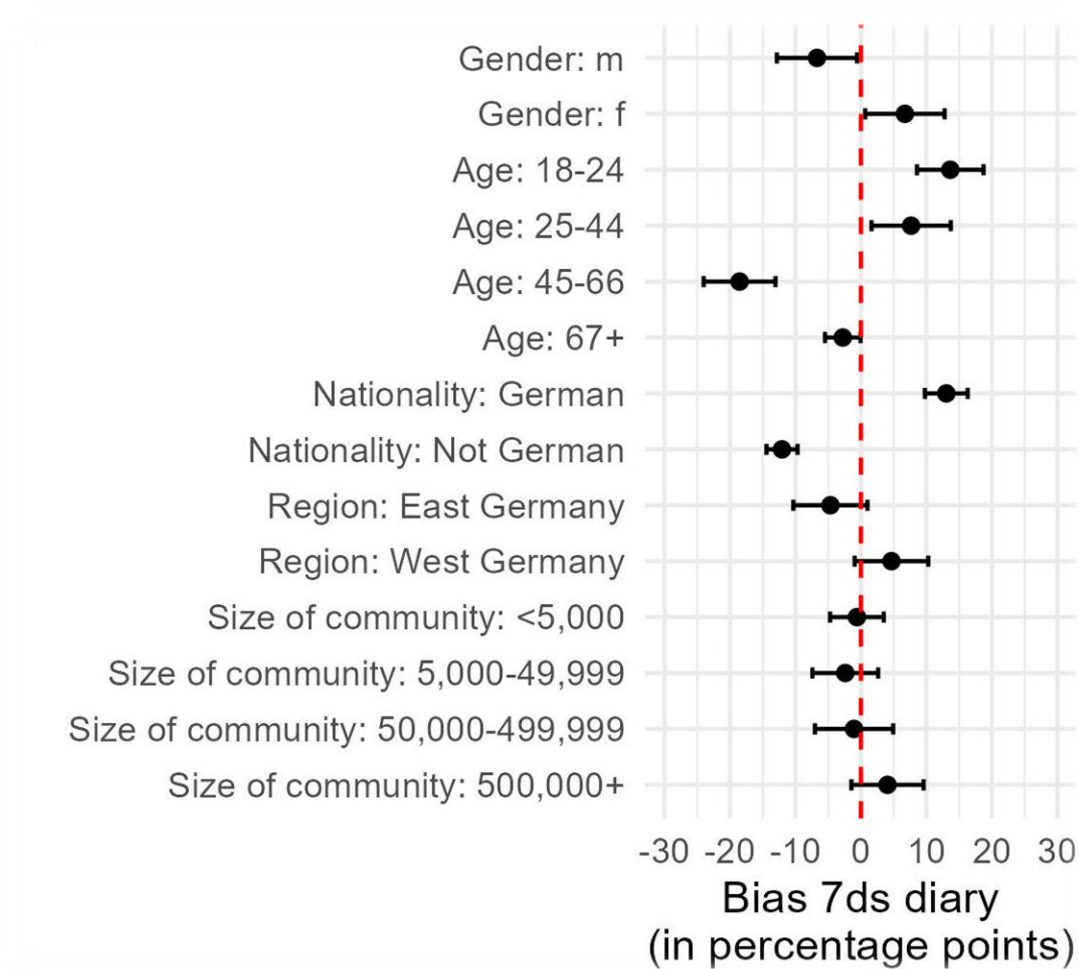
Results: Recruitment Experiments

- Belgium
 - Letter design: n.s.
 - Mode flexibility in follow-up: n.s.
- Germany
 - Letter stressing effort reduces download rate
 - Mentioning camera function: n.s.
- Norway
 - CATI interviewers increase starting and finishing rate
 - Type of download link: n.s.

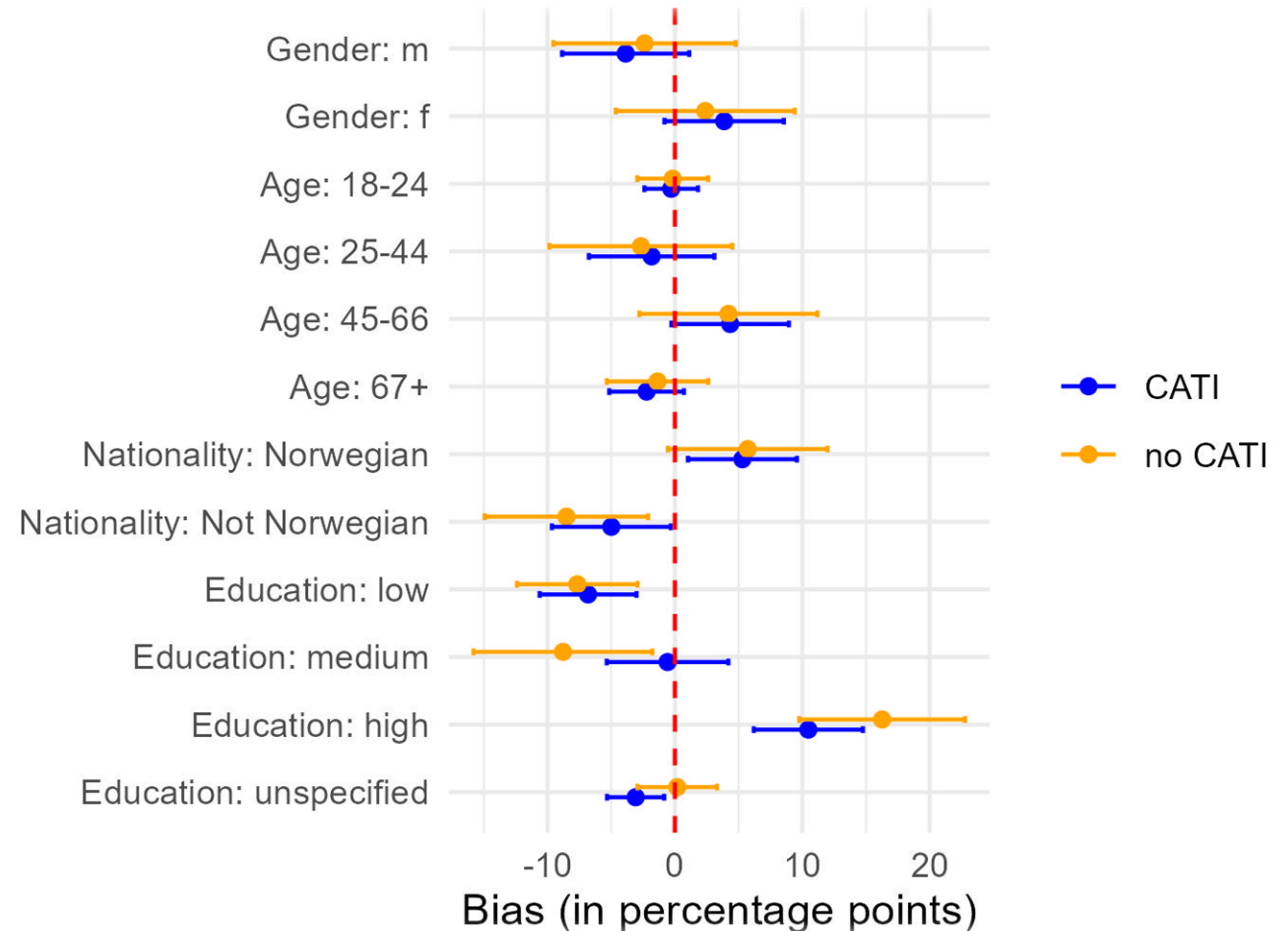
Results: Participation Bias in Belgium



Results: Participation Bias in Germany



Results: Participation Bias in Norway



Large Field Tests: Discussion & Conclusion

- Recruitment for smart surveys extremely difficult if no interviewers used
 - Use of CATI interviewers significantly increases the starting and finishing rate in Norway
- Relatively consistent non-participation bias regarding gender, age, nationality, and education across all three countries
 - CATI reduces education bias in Norway
- Manipulations in the invitation letters had very little effect
 - Stressing effort in invitations should be avoided (Germany)

Methodology and usability

- Quality of images not a big problem
 - Respondents can take pictures
 - Common issues like wrinkles, lack of contrast not a big problem
- OCR works well
- Which lines on receipt are relevant? an issue
 - See practical examples
 - Need for training of AI model so it recognizes the structure per shop
 - +- 300 receipts per big shop
 - Focus on supermarkets in project

Practical issue example: discounts



Nutri-boost discount milk

Nutri-boost discount on bread

Nutri-boost discount on peaches

Nutri-boost discount on broccoli

Total Nutri-boost discount on ticket...

...included as expenditure

KORTING NUTRI-BOOST

Name of expense item
KORTING NUTRI-BOOST

Type of expense item
tentative

Return or expense item	Price	Discount	Quantity	Price
return item	1.37	0.00 €	1	1.37
Keep item private	€			€
discard item from household budget				

Practical issue example: VAT

colruyt laagste prijzen

colruyt Food Retail B.V.
Oudegrasweg 108 • 1032 NC Amsterdam
Tel: 020 246 1000 • www.colruyt.nl

Tuuk ergens een lagere prijs gezien? Bel ons: 020 246 1000.
We passen onze prijs aan en betalen u het verschil terug.

276.2 L

Een smakelijke BBQ voor de laagste prijzen
Vind inspiratie, recepten, tips en heerlijke ingrediënten
voor al je barbecues op
colruyt.be/bbqbarbecue.

Zaterdag: 20/06/2024 11:55 Tel. winkel: 020 257175 Openingsturen: Uw kasabehouder:
CASH WILKOM Tel. bestelcentraal: 020 498576 Ma-Do: 08.30 - 20.00 Binn
Tel. winkel: 020 257175 Vrij: 09.00 - 21.00

Factuur

Art. Nr.	Benaming	Laagste	Howe	Quantiteit	Bedrag excl. BTW	Bedrag incl. BTW
A 11727	BONI ECO Koffie geselen Peru PT 250g		4	3,953	15,812	
A 7140	LIPTON Green Tea Intense Mint 20st		3	1,877	3,755	
A 11879	LIPTON Black Tea Forest Fruits 20st		3	2,18	4,321	
A 22298	LIPTON Mint Infusion 20st		3	3,085	6,17	
A 22287	LIPTON Infusion Camomille 20st		3	3,059	6,019	
A 22256	LIPTON Infusion Rosehip 20st		3	2,18	4,321	
Dranken						
C 21729	BREF badkamerreiniger cit.spray 750ml		3	3,38	6,199	
C 14674	BONI ECO ruitenspray 750ml		3	1,875	3,86	
C 17032	NIVEA deo Fresh Natural 200ml		3	3,199	6,199	
C 1311	NATURE BOX de avocado/zingen 368ml		4	3,626	15,308	
Professional Plus korting						
Non Food						
Totaal binnen Professional Plus kaart: € 8,58						
BTWCode A 9,00% C 21,00% Totaal						
Dranken						
Non Food						
Excl. BTW						
BTW						
Totaal korting leverancier (buiten BTW): € 8,58						
Totaal € 75,44						

BTW NO: € 42,82 BTW BEDIENING: € 31,98 BTW LEVERANCIER: € 0,10 BTW BETALING: € 75,44

007 291 943 0007 3524 EUROCARD/MASTERCARD 5 75,44

PAKID 8008 1801208 5 75,44

Tegemen 5 0,00

Personal data blacked out

3-digit prices are recognized as thousands

Total sum of 66.39€ (not shown on receipt) is calculated without VAT in microservice

Respondent needs to add VAT manually

Name of expense item

btw

Type of expense item

0000001 - Free addition / I don't know

Return or expense item

	Price	Discount	Quantity	Price
return item	9.05 €	0.00 €	1	9.05 €

Keep item private

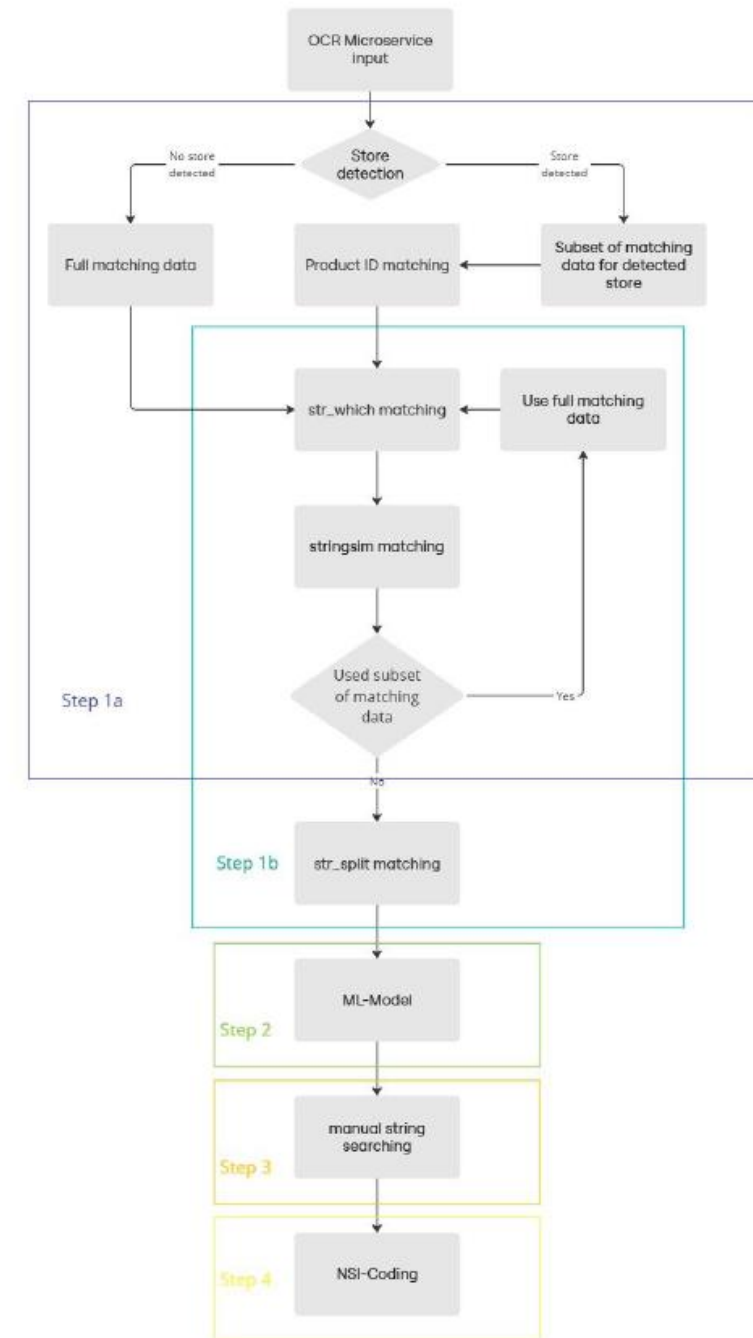
discard item from household budget

CoiCop matching

- The texts extracted from receipt need to be matched to a product category
- E.g. “milch 1 liter ProB” = “fresh milk”

How to do this?

- String matching
- Problems with new/changing products
- Good option do to this all in post-processing



Usability testing – Receipt Scanning

Respondents expect

- a receipt scanning microservice
- processing time to be fast
- processed data to be very accurate

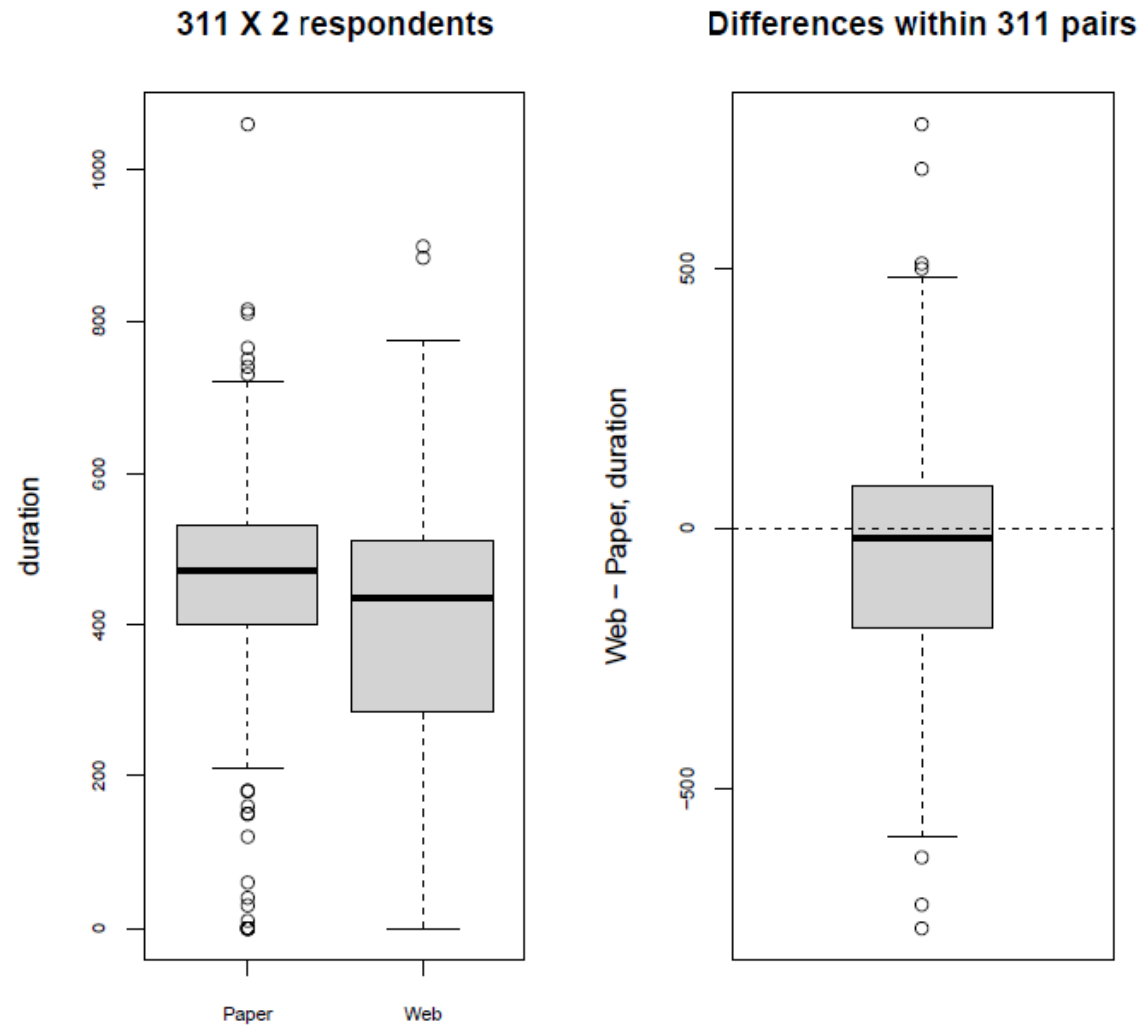
Respondents do not expect

- to commit (correct data)
- to classify expenses
- to provide additional context
- to read many notifications

Future challenges lie with

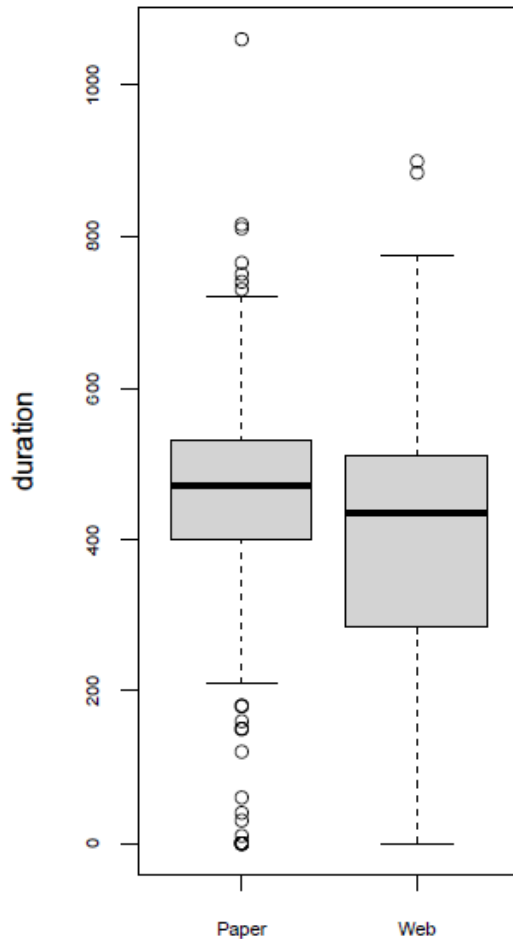
- training microservice
- dealing with steep learning curve
- communication with respondents
- adding and training classification algorithms
- alternatives for paper receipts (digital)

France: TUS differences in total sleep duration

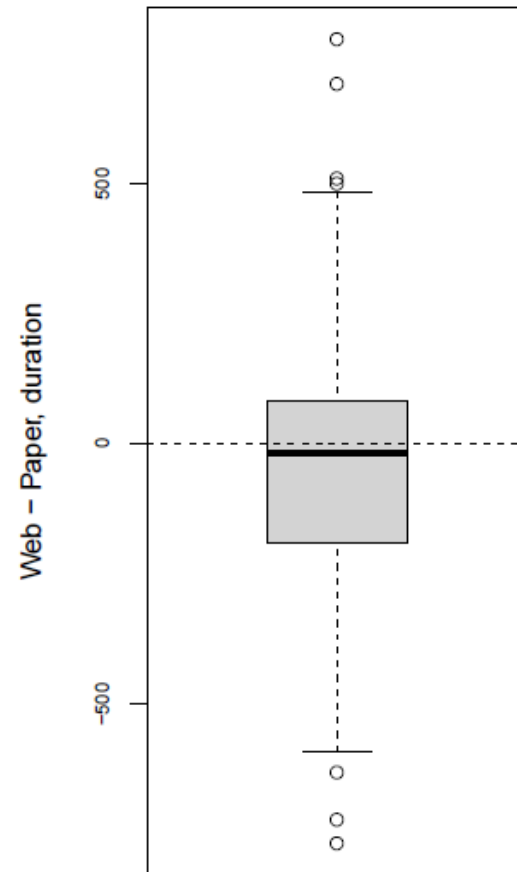


TUS differences in total sleep duration

311 X 2 respondents



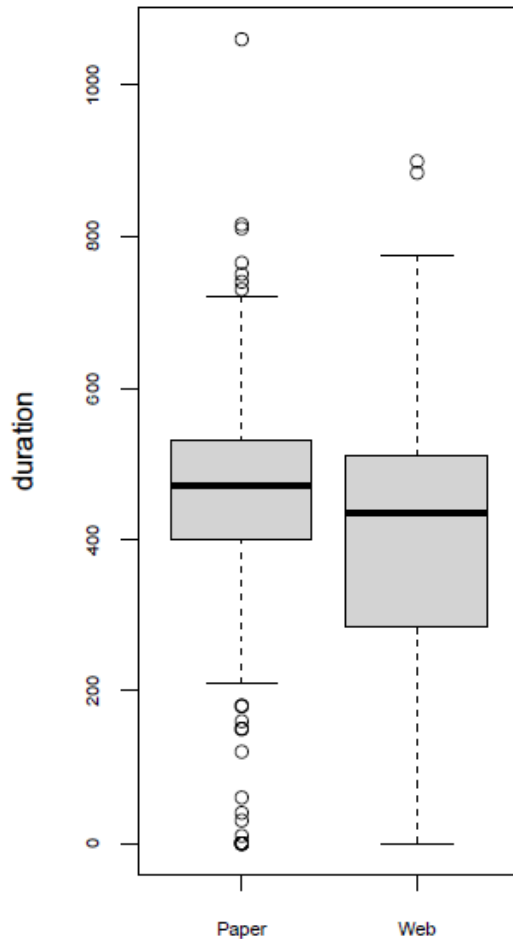
Differences within 311 pairs



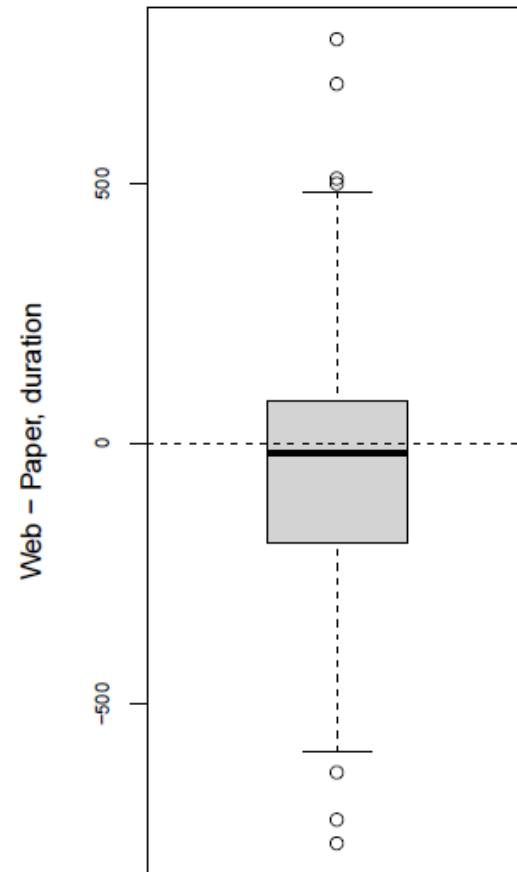
- 28% of internet-responders < 5hrs sleep between 4am and 4am
- 14% of paper-responders < 5hrs sleep between 4am and 4am

TUS differences in total sleep duration

311 X 2 respondents



Differences within 311 pairs



- 28% of internet-responders < 5hrs sleep between 4am and 4am
- 14% of paper-responders < 5hrs sleep between 4am and 4am

Possible explanations

- Mode differences in missing data
- Interviewer effects
- Can be corrected in post-processing

TUS differences in other duration

	Crossover-based constant effect	e.g. “on web more time on sociability”
Personal and physiological time	-1 [-20; 18]	
Working and studying	-8[-24; 8]	
Housework and recreation	-7 [-16; 3]	
Care	4 [0; 8]	
Sociability	25 [17; 33] *	
Leisure	-51 [-64; -38] *	
Travelling	-20 [-27; -14] *	
Note: * $p < .05$		

Overall conclusions

SSI Work Package 2	Task 1	Recruitment to smart surveys very difficult without interviewers Bias along lines of traditional digital divide
	Task 2	ML and processing works with good input data (receipts or GPS data) Need for country-specific model/training (HBS)
	Task 3	Smart microservices also needs to be clever Smart microservices lower respondent burden but not complexity of survey
	Task 4	Large measurement differences between modes Likely a combination of causes