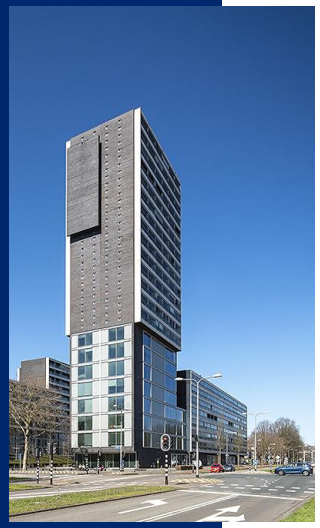
Can You Hear Me, Loud and Clear?

Advantages and Limitations of Voice Recorded Speech to Text Answers in the Online LISS panel

Joris Mulder
LISS coordinator
senior researcher

Speech ASR seminar SURF Utrecht

25 June 2024



Independent non-profit research institute for **data collection** and **applied research** at the campus of Tilburg University

Currently ~ 50 colleagues, plus a number of student assistants

Centerdata mainly works for

- the academic community
- policy makers / government institutions
- European Commission and EU surveys ESS, SHARE, EVS

Centerdata's core activities

Survey Research

Online (panel) surveys
Target groups
Experiments
Data dissemination

IT

Software Development

Policy Research & Analytics

Forecast Labour Market (Education)

Data Science

Machine learning
Deep learning
Text analytics
Data maturity
Data visualisation

Consumer Research

Behavioural Research

What is the LISS panel?

- Online longitudinal research panel on household level
- 7,500 panel members (16+) from 5,000 households
- Operational since 2007

Online surveys and experiments

- Panel members complete surveys every month
- About 60 minutes survey time per month
- Incentive of € 2.50 for 10 minutes completion time

Representative

- Probability based samples, drawn bij Statistics Netherlands (CBS)
- Loyal panel members: high response (70-80%), low attrition (10% yearly)
- Non-internet households included → simPC
- Scientific, policy or socially relevant research: non-commercial

Additional unique features

- Panel management department: help and technical support
- All collected data in the LISS Data Archive
- LISS data can be merged with CBS microdata

Recruitment: sequential multi-mode strategy

Letter and brochure

B1 language level

5 euro pre-paid incentive

Unconditional

Online and classical approach

CAWI – CATI – CAPI,
follow-up CAPI en PAPI

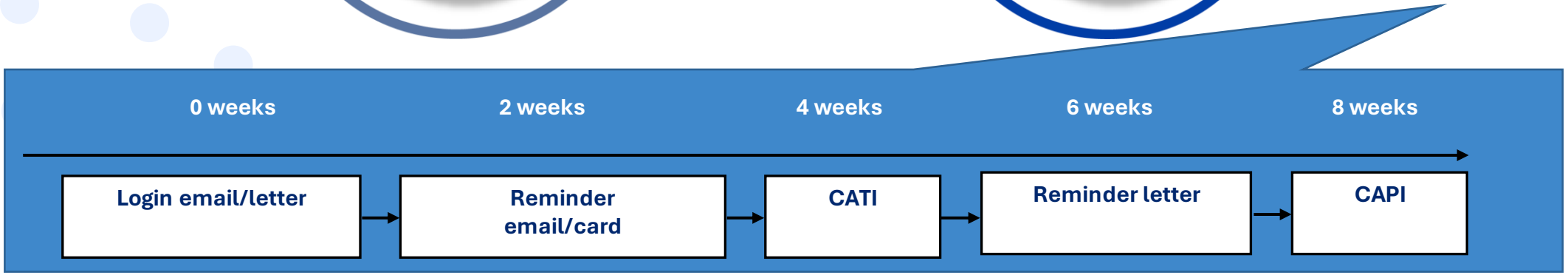
Recruitment interview

About 10 minutes

DSL and simPC

10 euro post-paid incentive

Conditional: 10 euro post-paid after registrering



Non-response patterns

- **Similar**

  to those of other leading scientific panels

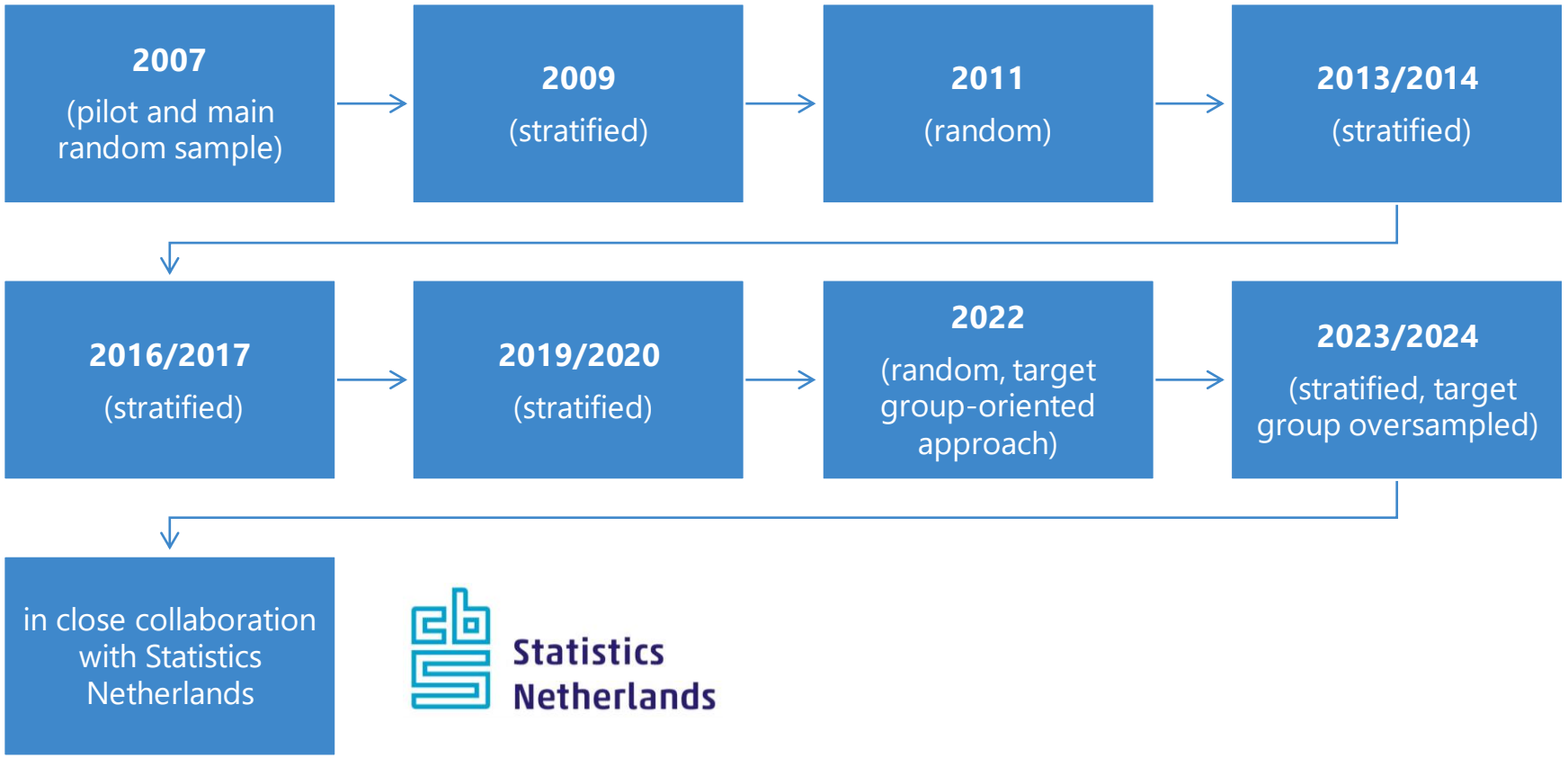
- **Superior**

  to commercial access and volunteer panels

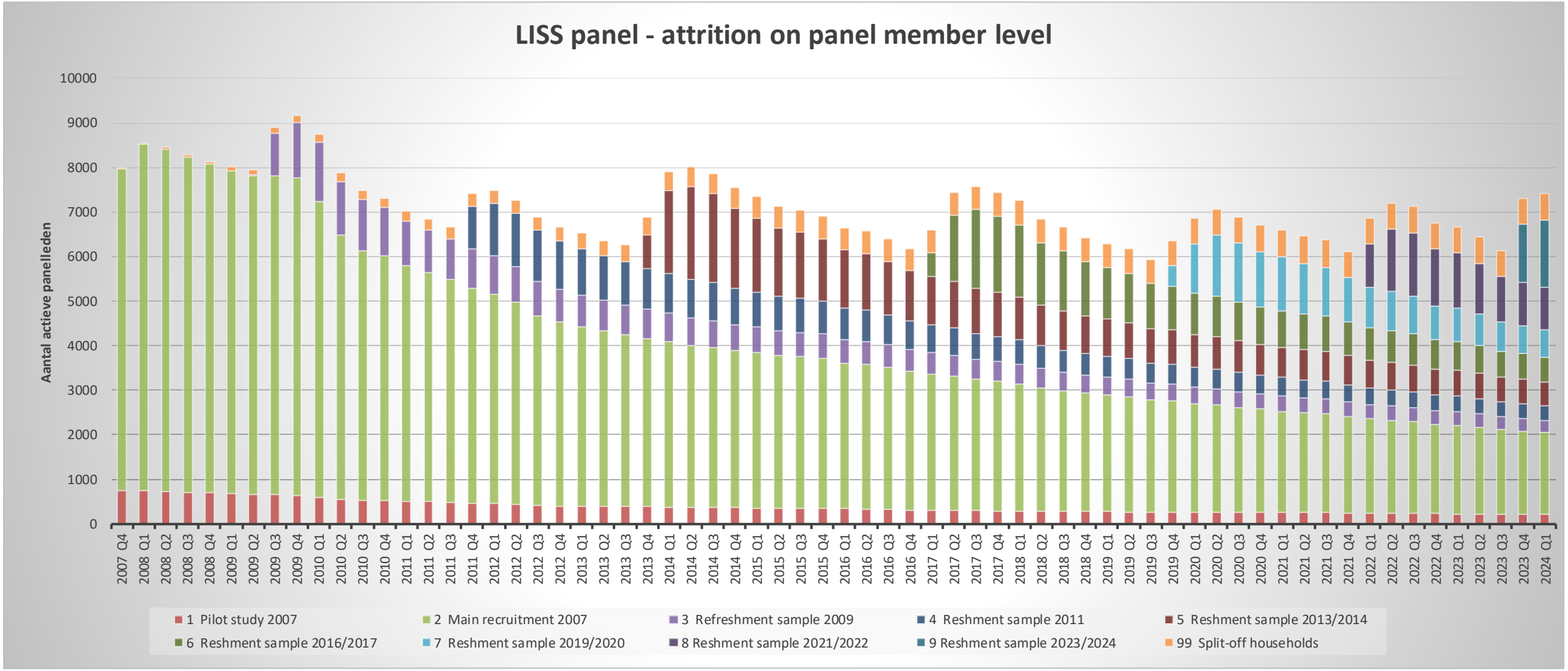
✓ no coverage problems

✓ no self-selection

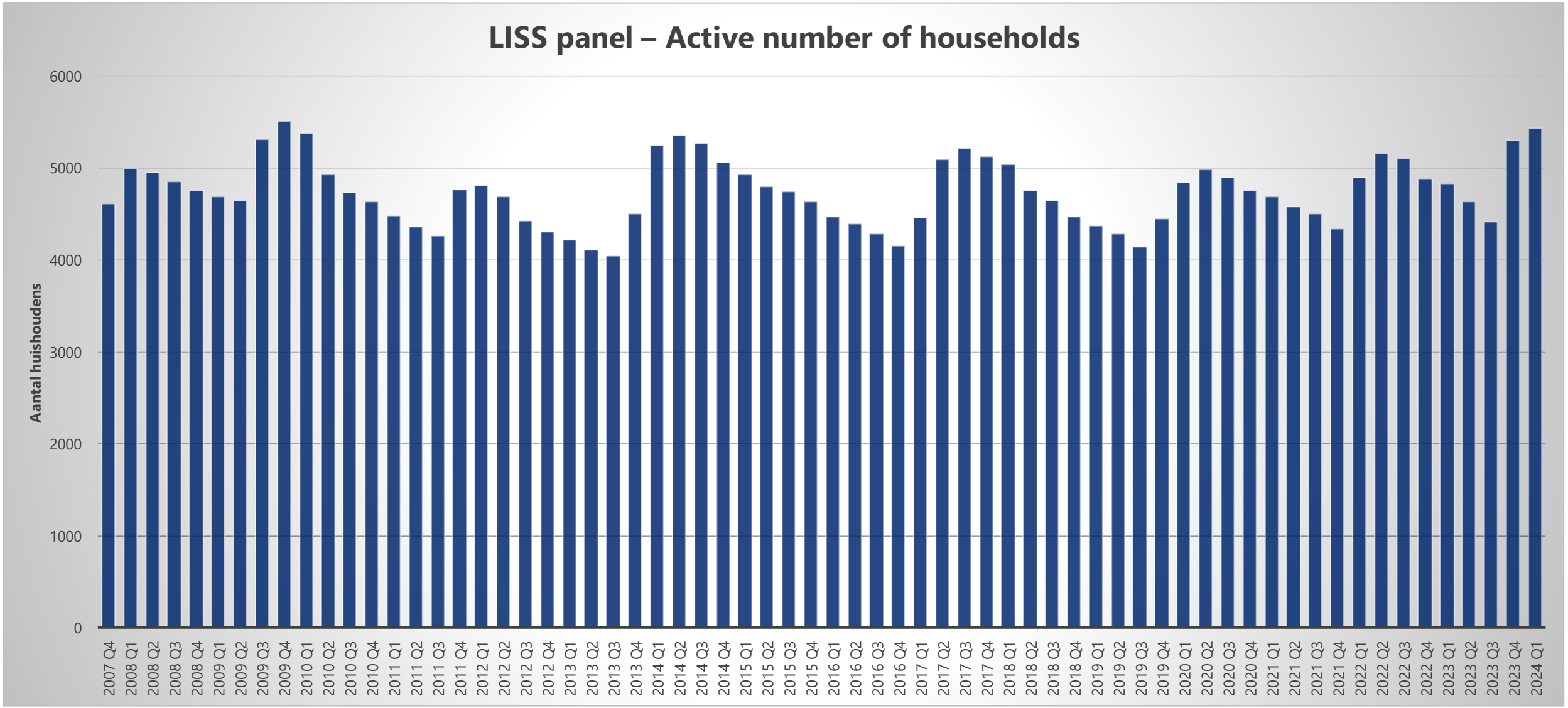
Refreshment samples



Attrition and refreshment (1)



Attrition and refreshment (2)

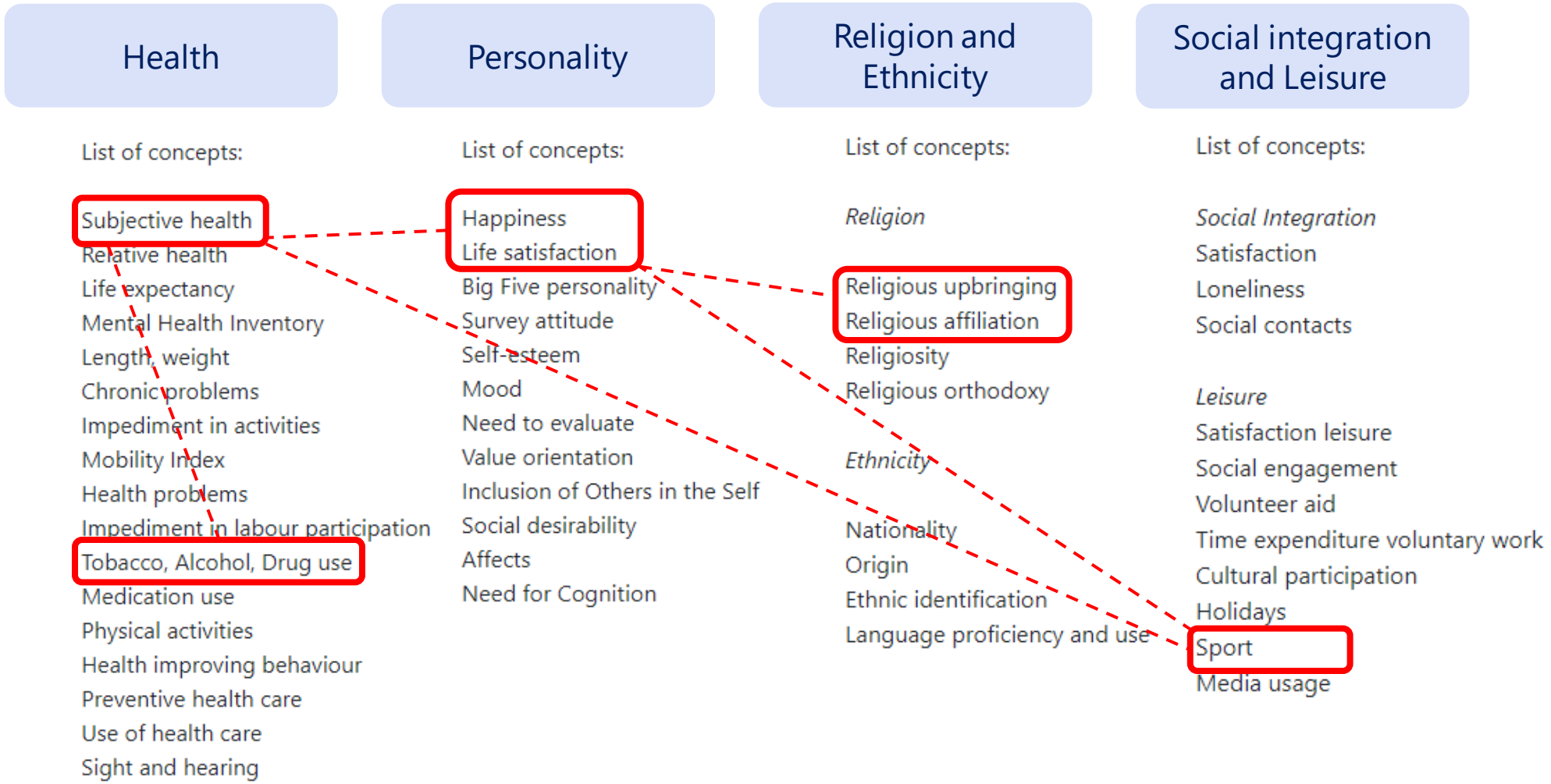


Use of the LISS infrastructure

1. Collection of **new** data
   - proposals can be submitted throughout the year
     - budget required
   - Calls for free use, through ODISSEI

2. Use of **existing** data
   - **free of charge**
   - longitudinal core study and proposed studies, since 2007

Example of linking Core Study data



The LISS panel

Longitudinal Core Study

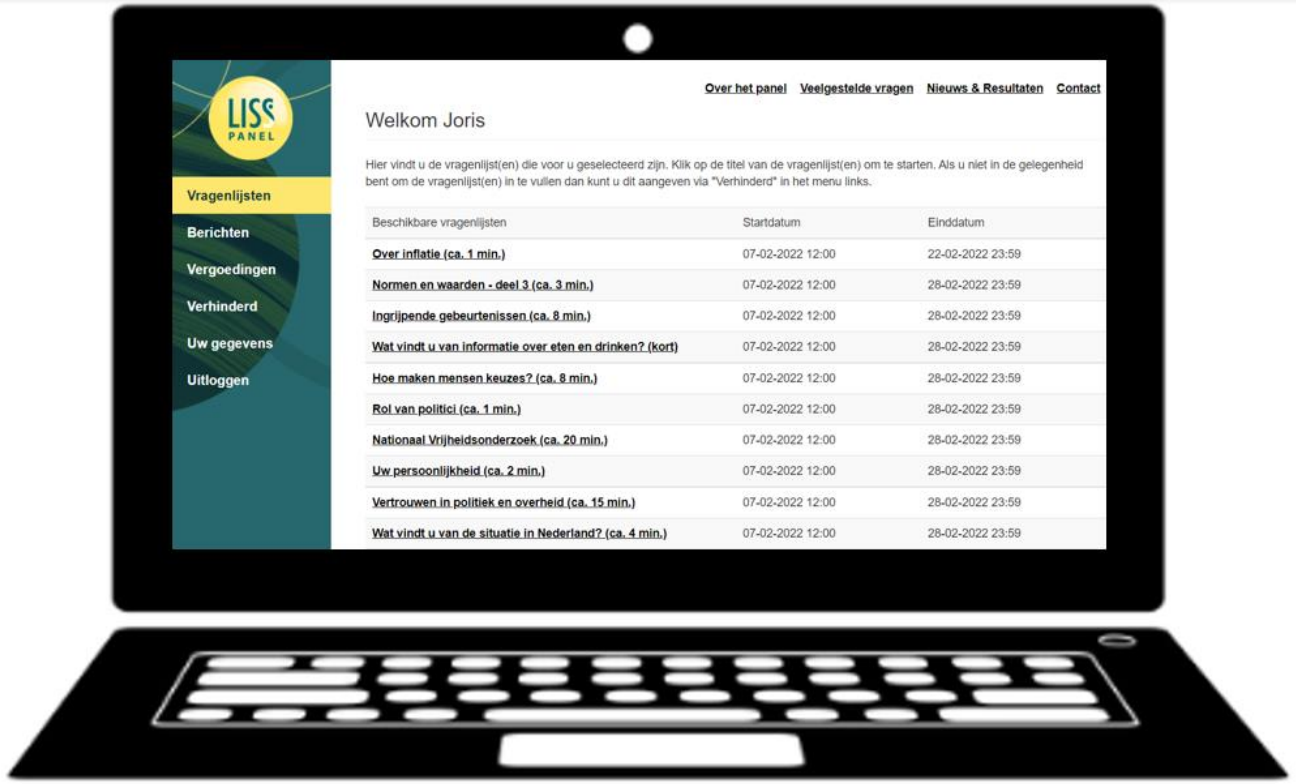
Surveys and experiments

Innovative studies

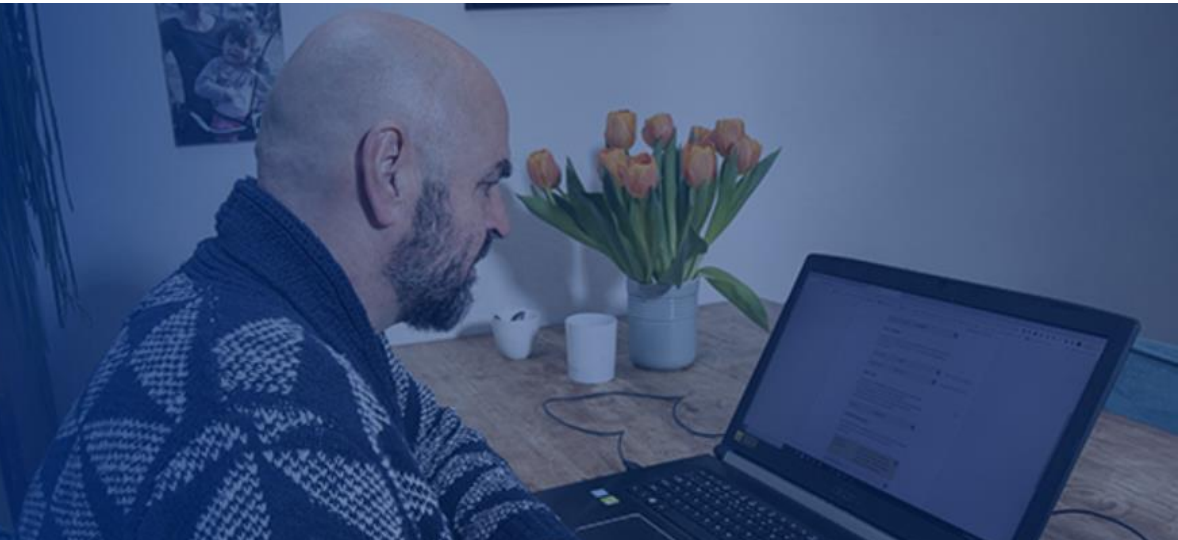
Microdata Statistics Netherlands (CBS)

LISS panel

Statistics Netherlands



High-quality online research infrastructure for the social sciences, available to researchers worldwide



The ODISSEI LISS CfP

- Longitudinal Core Study
- Surveys and experiments
- Innovative studies
- Microdata Statistics Netherlands (CBS)

LISS panel

ODISSEI
Open Data Infrastructure for Social Science and Economic Innovations

Statistics Netherlands

ODISSEI

- **~45 member organizations**
- Centerdata
- Central planning bureau
- NSCR, NIDI, PBL, SCP
- Dutch universities
- Nivel, eScience Center
- Dutch Central Bank

ODISSEI LISS Call for Proposals

https://odissei-data.nl/en/

Link to CBS microdata

Longitudinal Core Study

Surveys and experiments

Innovative studies

Microdata Statistics Netherlands (CBS)

LISS panel

Statistics Netherlands

See website CBS (for now in Dutch, but contact LISS if you need help):
https://www.cbs.nl/nl-nl/onze-diensten/maatwerk-en-microdata/microdata-zelf-onderzoek-doen/microdatabestanden/lisspanel-procedure-lisspanel-met-cbs-microdata

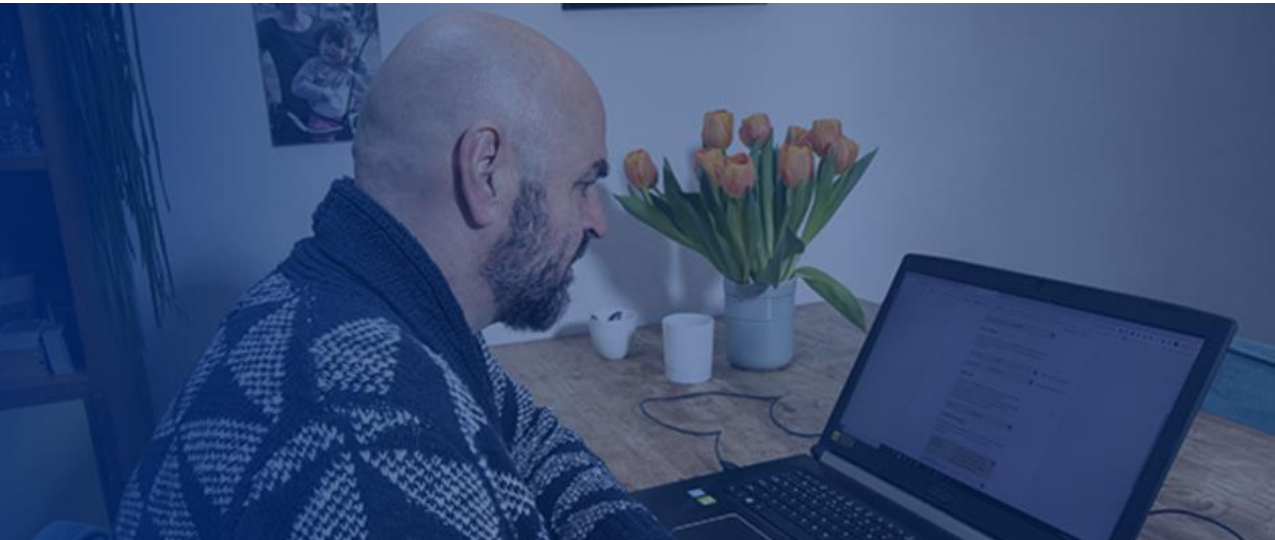
All data collected in the LISS panel can be linked to register data available at Statistics Netherlands (CBS, Remote Access)

Already used in studies on income, assets, and pensions, e.g.

De Bresser, J. and M. Knoef (2015). Can the Dutch meet their own retirement expenditure goals?, Labour Economics, 34, 100-117.

Zimpelmann, C. (2021). Stock Market Beliefs and Portfolio Choice in the General Population, Discussion Paper Series – CRC TR 224, DP No. 258, University of Bonn and University of Mannheim.

Other sources: e.g., weather data (KNMI), air pollution (RIVM)



Annual open ODISSEI LISS calls

ODISSEI
Open Data Infrastructure for Social Science and Economic Innovations

Microdata Access Grant (MAG) 2024 call guidelines

Apply for the MAG

2024 LISS grant call guidelines

Apply for the LISS grant call

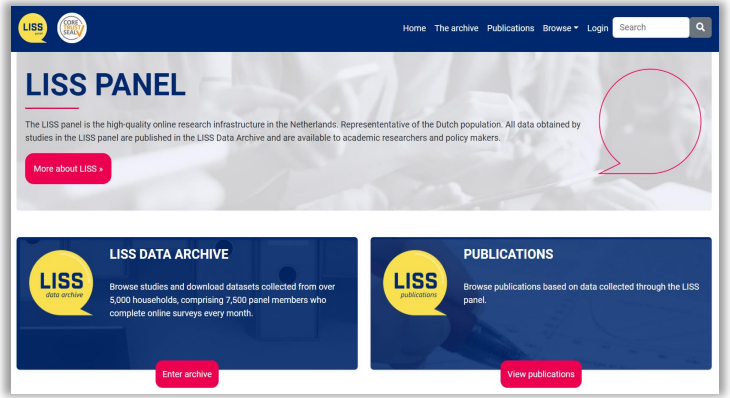
Since 2018

- Yearly 30-40 proposals
- 42 proposals granted
- 2 ad-hoc Corona proposals

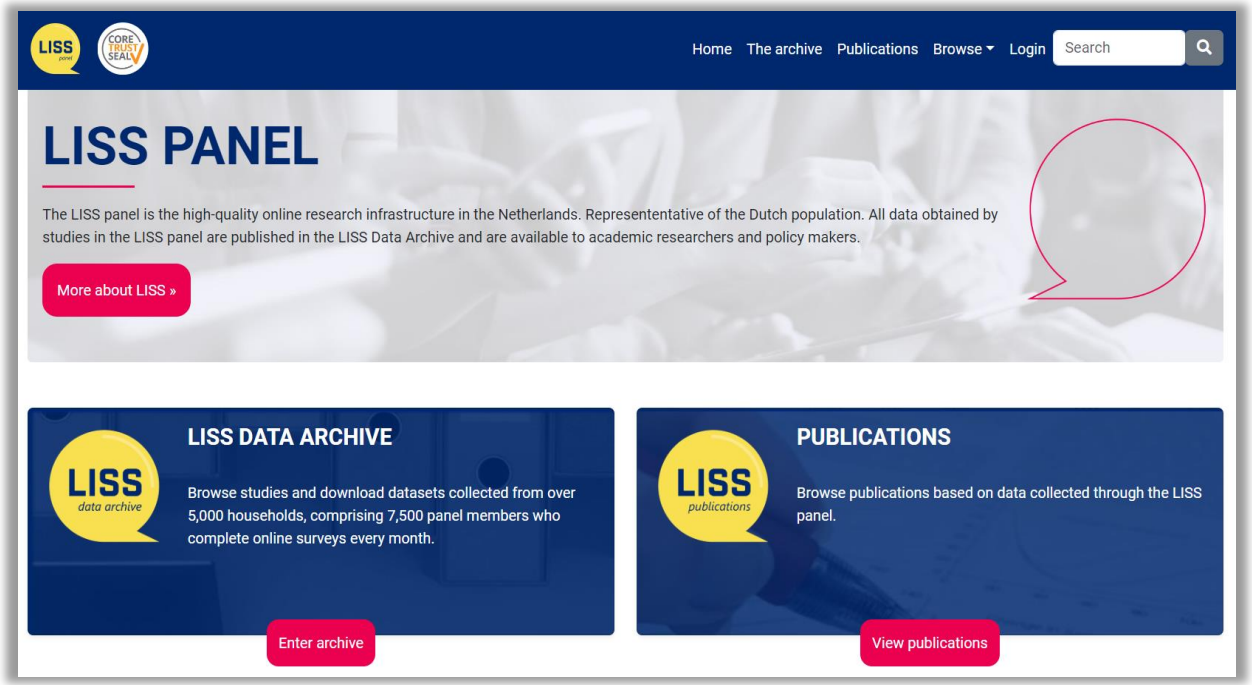
Examples

https://odissei-data.nl/en/category/research/

Lissdata.nl



LISS Data Archive



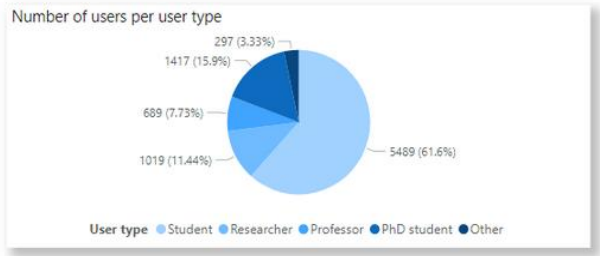
All data are easily available at no cost through the LISS Data Archive: https://lissdata.nl

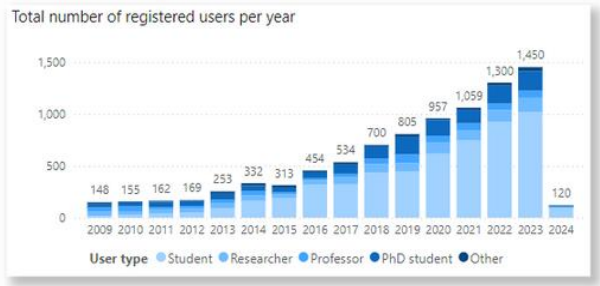
- More than 8,000 researchers
- Over a 1,200 publications based on LISS data
- Including about 700 articles in peer-reviewed journals and over 60 Ph.D. theses

LISS data archive - user statistics

8,911 Number of users

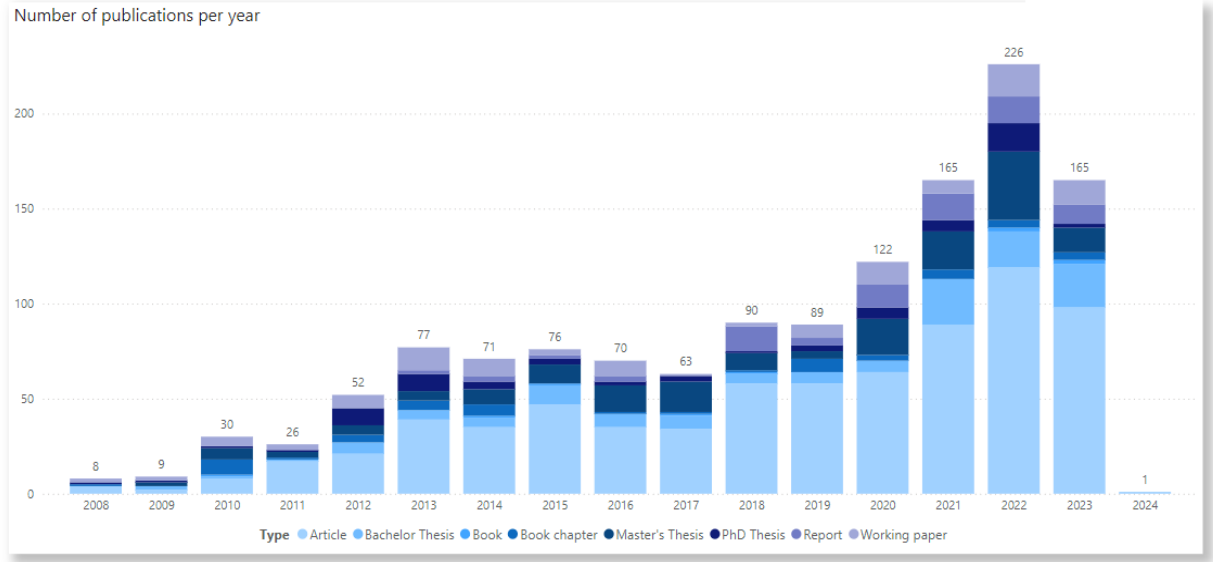
7,247 Number of active users





LISS data archive - publication statistics

1,340 Number of publications



Research in the LISS panel

Use of the infrastructure

Research in the LISS panel

Annual longitudinal LISS Core Study since 2007

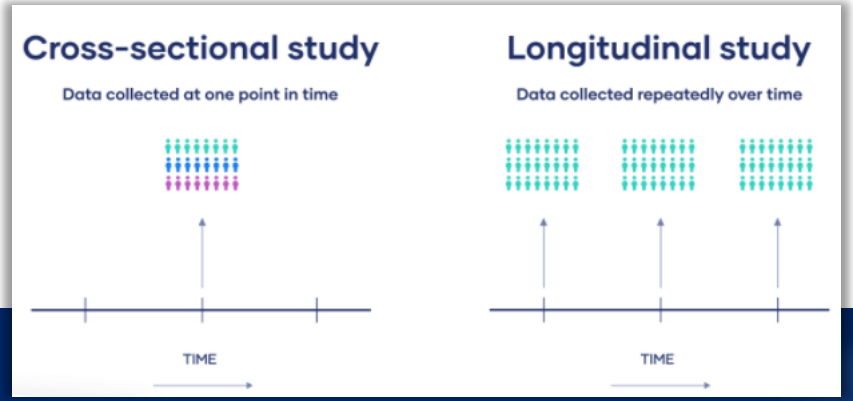
Under Centerdata management

- 2 Health
- 3 Religion and Ethnicity
- 4 Social Integration and Leisure
- 5 Family and Household
- 6 Work and Schooling
- 7 Personality
- 8 Politics and Values
- 9 Economic Situation: Assets
- 10 Economic Situation: Income
- 11 Economic Situation: Housing

Open for reseachers collecting new data

Budget required
ODISSEI LISS Call for Proposals

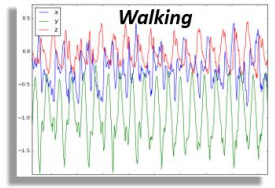
ODISSEI
Open Data Infrastructure for Social Science and Economic Innovations



Innovative studies









What is Speech to Text in online surveys?

Open-ended survey questions

- respondents answer by **voice**
- microphone (CARI)

Automated Speech Recognition (ASR)

- transcribing audio to text
- Storing raw audio files (optional)

- **Advantages** and **limitations**

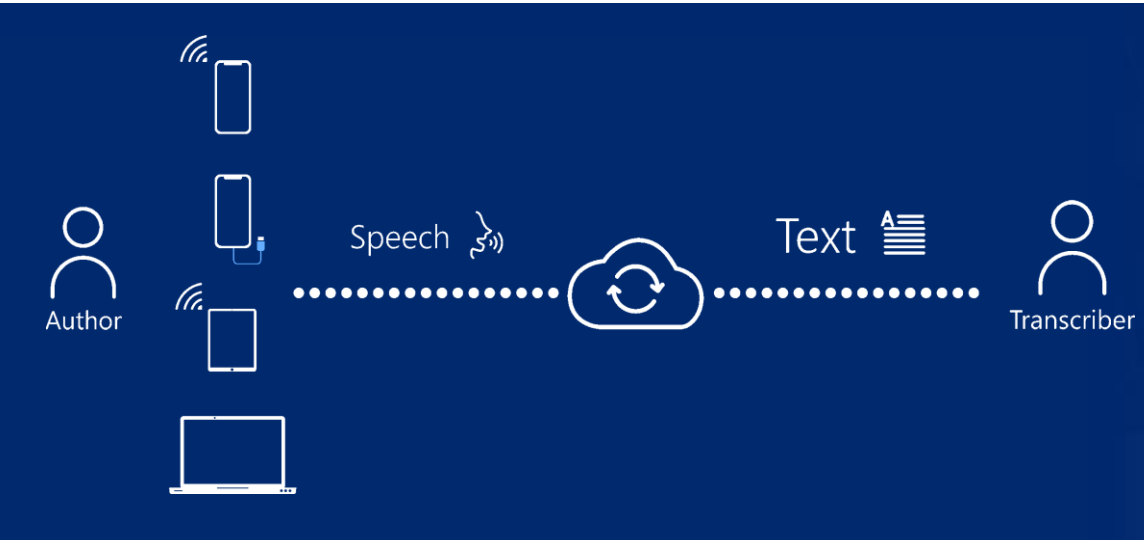
- two S2T studies in the LISS panel

1. **Randomized experiment** (Meitinger et al., 2024)
   - voice * text-response
   - respondent preference & experience
   - accuracy and validity of ASR

2. **Quasi experiment** (van den Heuvel et al., 2023)
   - voice * text-response
   - voice response > text-response
   - quality and usability of audio and ASR

Advantages

of Speech to Text in online surveys



Advantages

- Potential reduction of survey time (Revilla et al., 2020)
- Potential improvement of criterion validity (Gavras & Höhne, 2022)
- Automatic Speech Recognition (ASR) saves budget and time (Revilla and Couper 2021; Ziman et al. 2018)

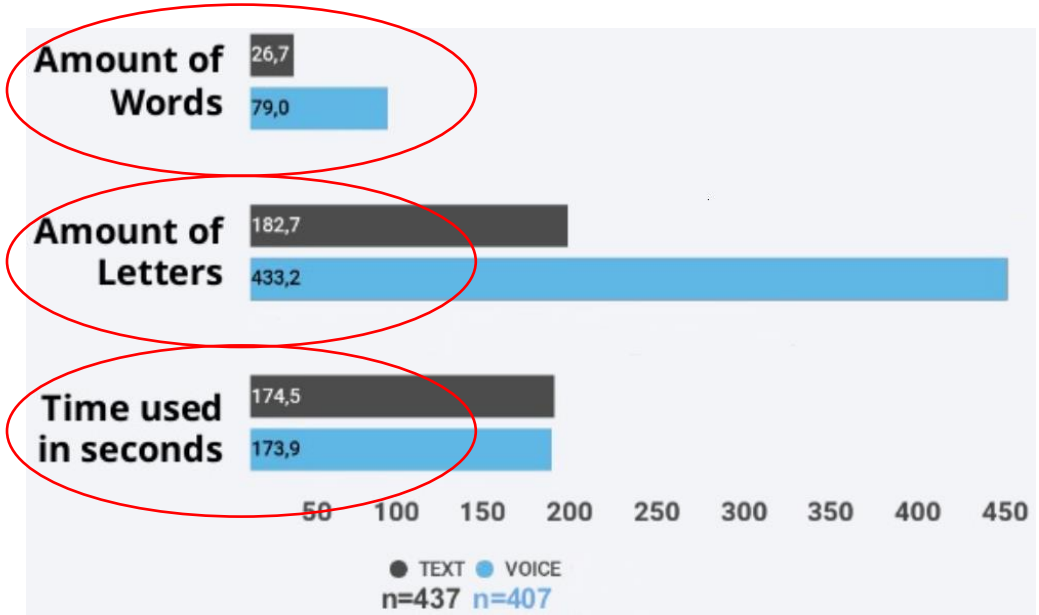


image by Questfox

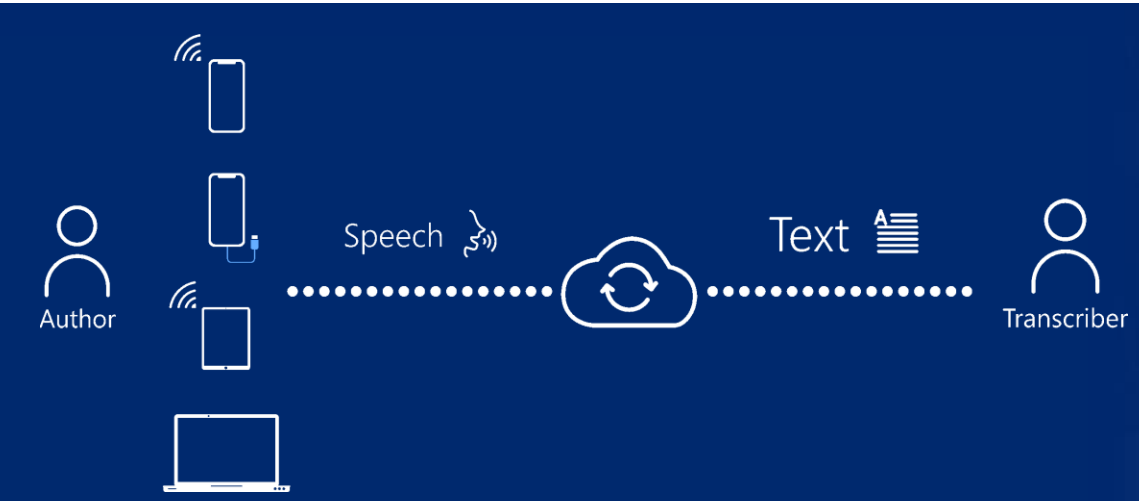
- Voice is valuable data source to measure
  - Cognitive functioning
  - Socioeconomic status
  - Verbal reasoning abilities
  - Emotion analyses

  (van den Heuvel et al., 2023)

Tone	NLP
Language proficiency	Topic modelling
Vocabulary	Sentiment analysis

Limitations

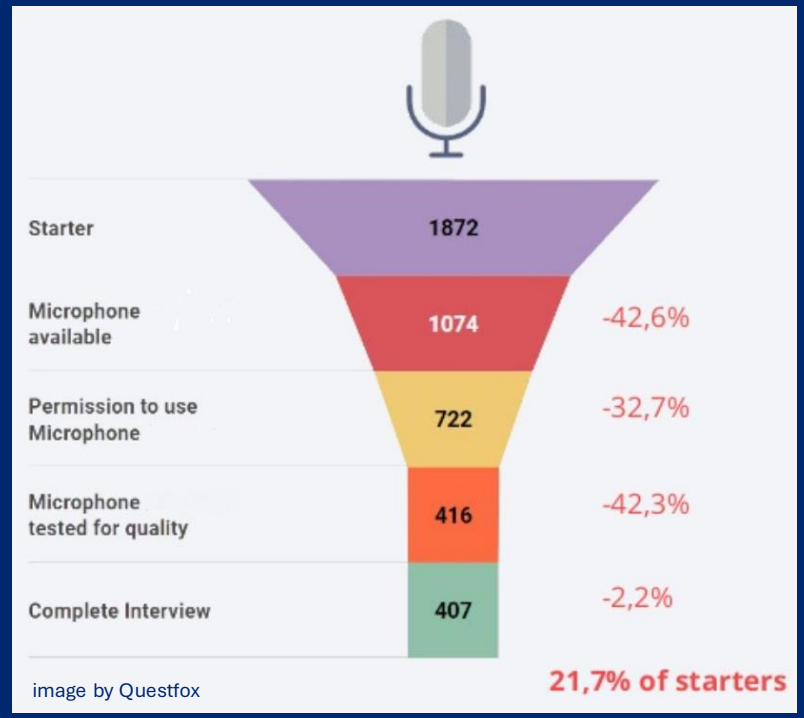
of Speech to Text in online surveys



Limitations (1)

- **Response decrease & bias**
  - Willingness to participate
  - Technological illiteracy
  - Technical constraints

- **Practical constraints**
  - Server load
  - Privacy and security
  - Integrate S2T in survey software
    - Technical integration
    - Respondent usability



Limitations (2)

- Manual audio transcription (conversion to text) costly and labor intense

- Automatic Speech Recognition (ASR)
  - **Accuracy** ASR can differ, due to longer, shorter, missing or added text (Errattahi et al. 2019; Ghannay, Estève, and Camelin 2020)
  - Word Error Rate (WER)
    - Number of errors divided by answer length (Kim et al. 2019; Tancoigne et al. 2022)
    - The higher the WER value, the worse the transcription

  - **Validity** ASR can change the meaning of transcribed words

ASR transcription example

“Wat eet u meestal tijdens de lunch?” (Dutch answer)

“What do you usually have for lunch?” (English translation)

	Dutch answer	English translation
Voice audio	Ik eet meestal een wortel	I usually eat a carrot
Transcription	Ik weet meestal een gordel	I usually know a seatbelt

For the sake of the argument…

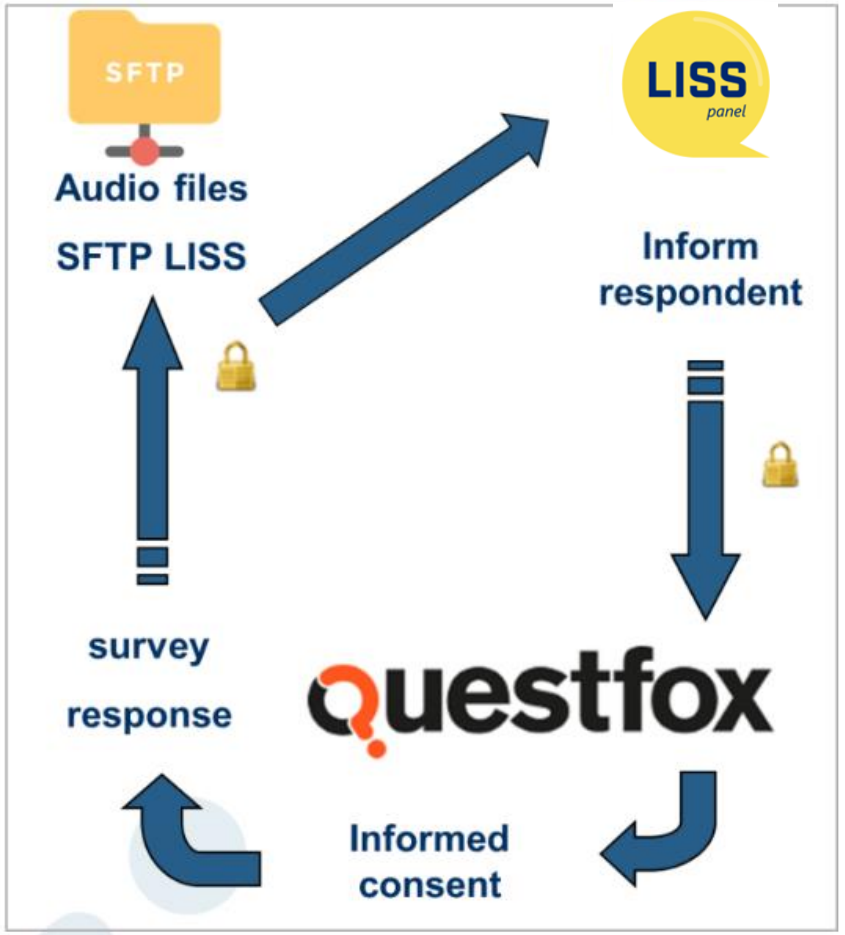
I usually eat a carrot
I usually eat a rabbit

Low accuracy (higher WER value)

→ Deteriorates validity (meaning) (Meitinger et al., 2024)

S2T Integration in the LISS panel – Questfox SaaS

LISS S2T flow logic



S2T survey example: https://youtu.be/EH1R6myXB-o

First Speech to Text experiment in the LISS panel

Meitinger et al., 2024

RQ1: Does the **accuracy** of ASR transcriptions differ by subgroups and context factors?

RQ2: Does the **validity** of ASR transcriptions differ by subgroups and context factors?

Subgroups: sex, age, education

Context factors: alone or not, background noise

What kind of questions?

In general, how would you rate the current state of the economy in the Netherlands?

1 Very good

2 Good

3 Not good, not bad

4 Bad

5 Very Bad

99 Don't know

Please explain why you selected [answer]

Keep the noise down: On the performance of automatic speech recognition of voice-recordings in web surveys

Katharina Meitinger, Sabien van der Sluis, Matthias Schonlau, 2024

Fielded in December 2020

- Experiment with 3 conditions
  - 5 min. survey
  - Track C: only n=88 chose voice!

- Overall 76% response

~ 20% screened out

~ 50% completed

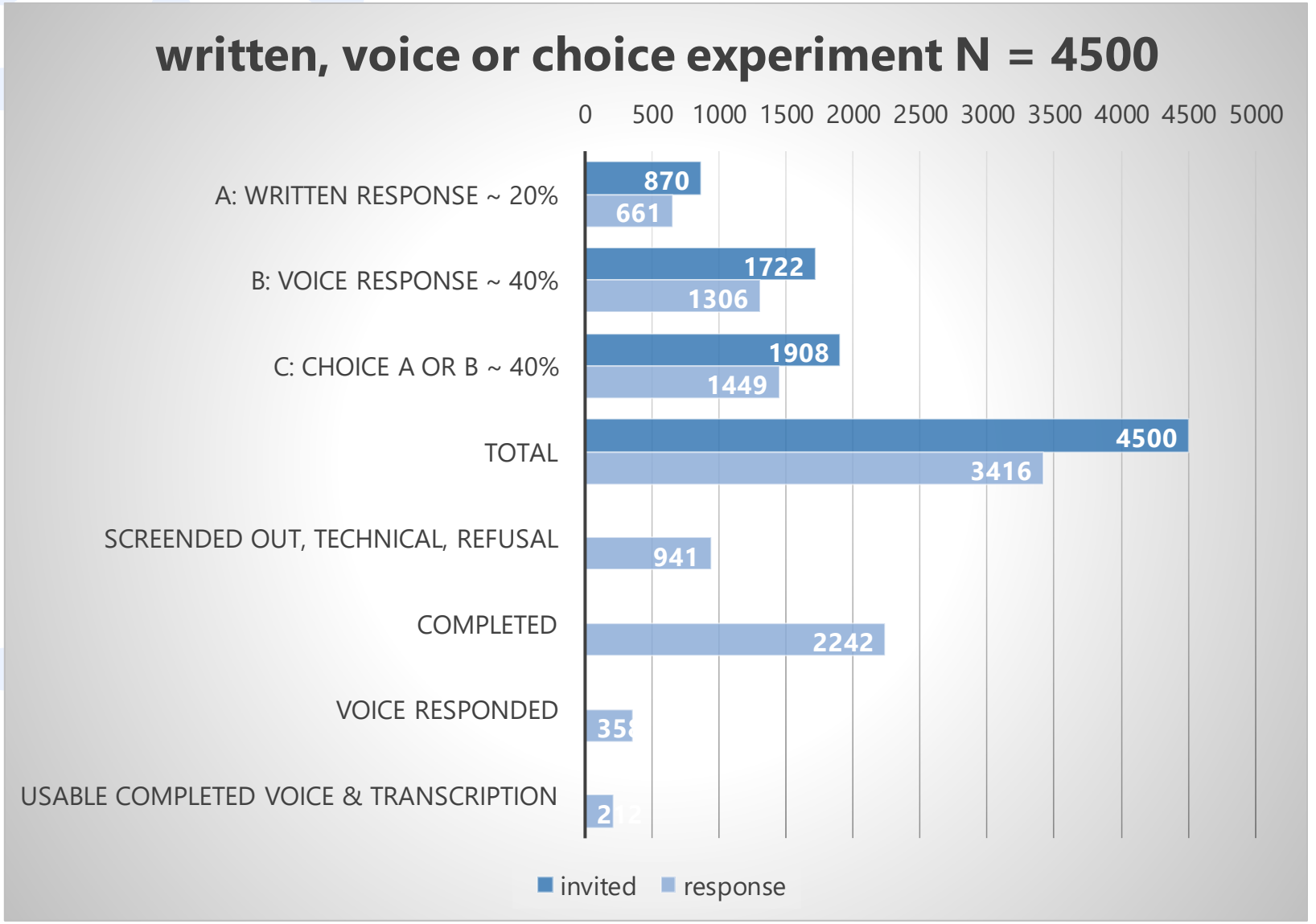
~ 8% voice response

~ 5% usable voice responses

- Collected audio files:

~ 1,430

~ 1,000 good quality

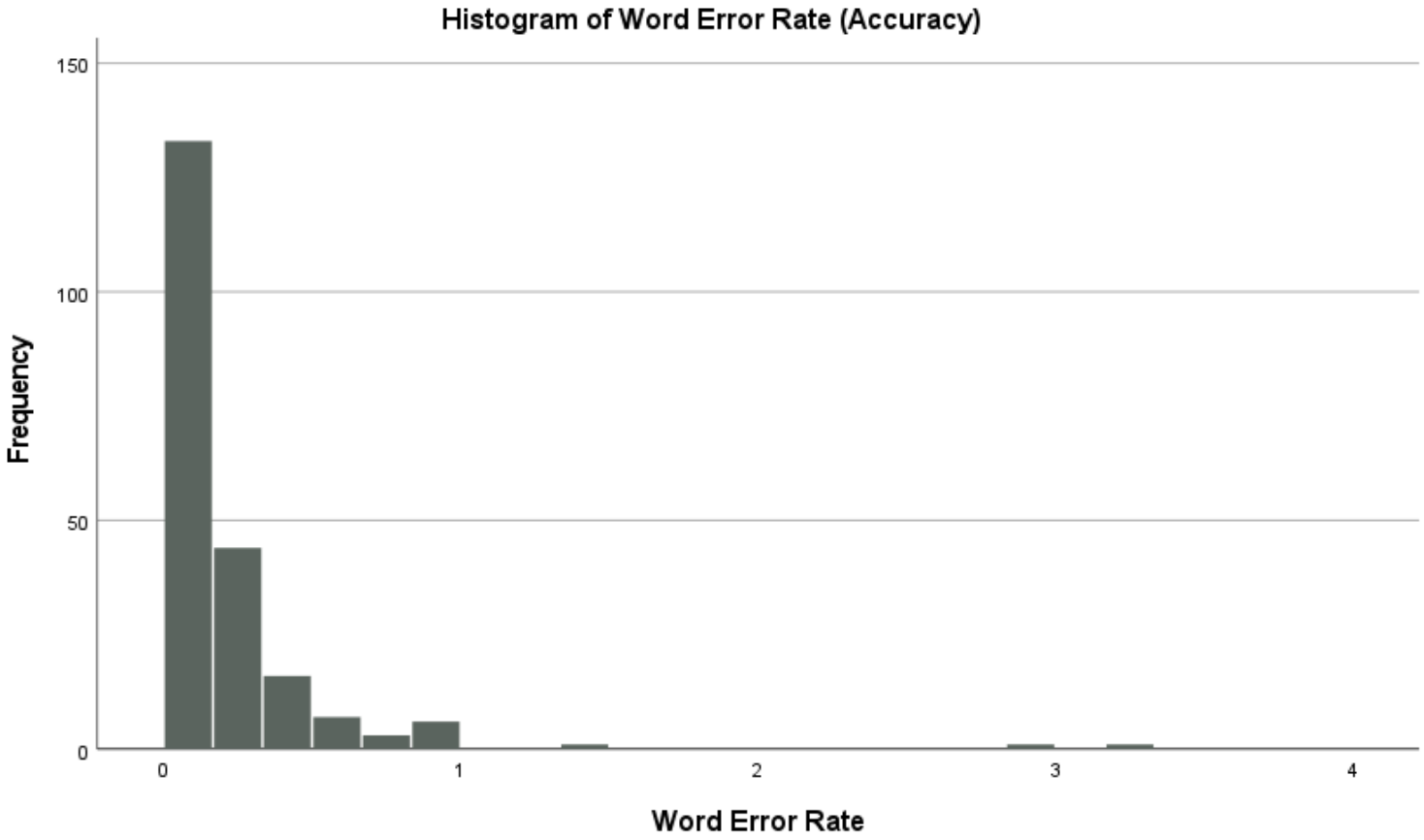


Accuracy

Word Error Rate (WER) ranged from 0 to 3.33

Average transcription WER was 0.20 (SD=0.36)

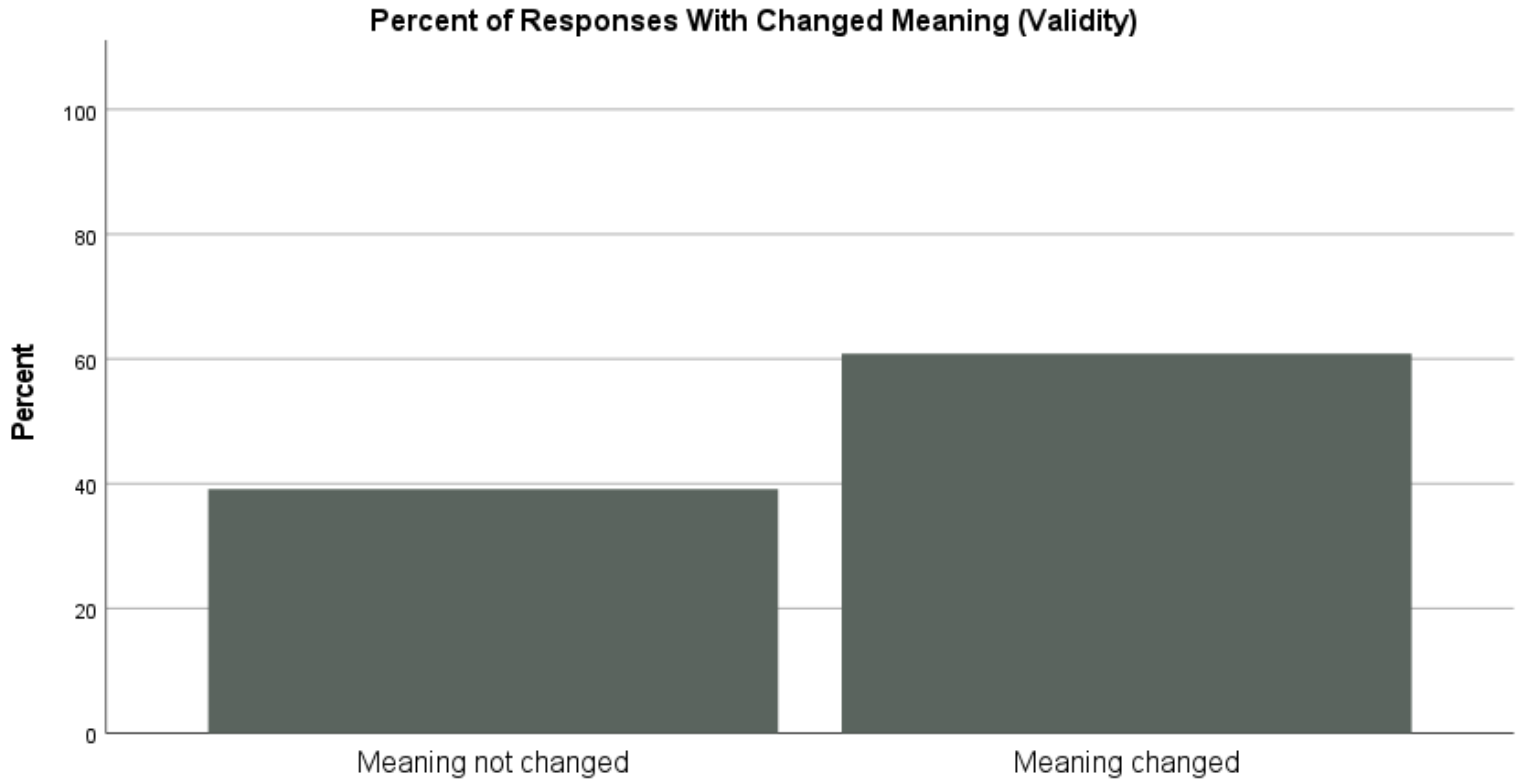
Which means that **20%** of the words would need to be **altered** (via substitutions, deletions, or insertions).



Validity

In **60.8%** of the analyzed responses, the **meaning** of at least one word **changed** due to the ASR transcription.

Responses with **background noise** had 2.21-times **higher odds** that the **meaning** of the response **changed** than responses without background noise (*p*=.030).

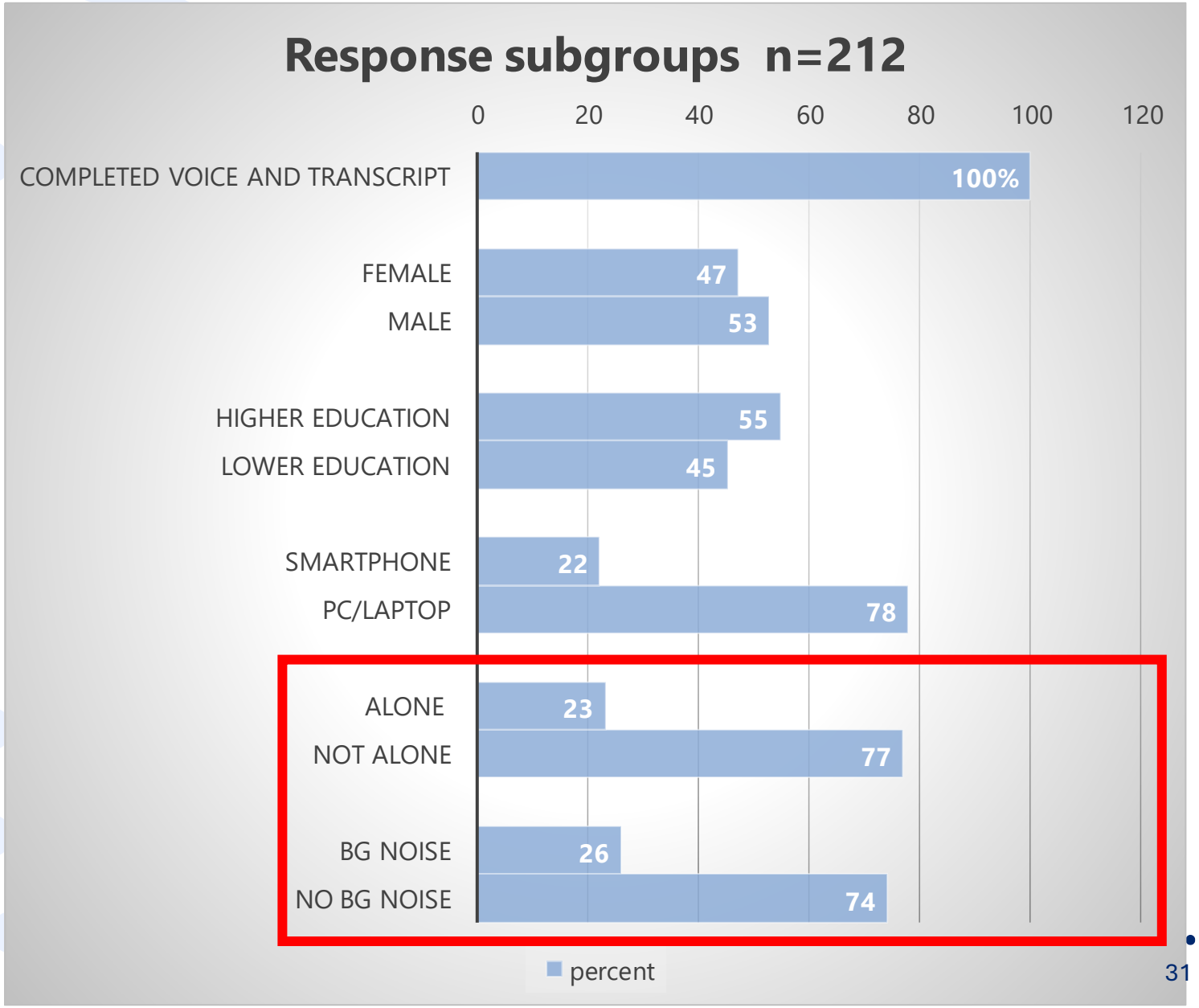


Main findings Meitinger et al., 2024

Background noise reduces accuracy and validity of ASR transcriptions.

Validity improved when respondent was alone vs not alone (OR: 0.43, p=.017).

No accuracy or validity differences across age, sex, education, device or location.



Second Speech to Text experiment in the LISS panel

What kind of questions?

15 open-ended questions.

What are the most important characteristics of a democracy according to you?

What does marriage mean to you?

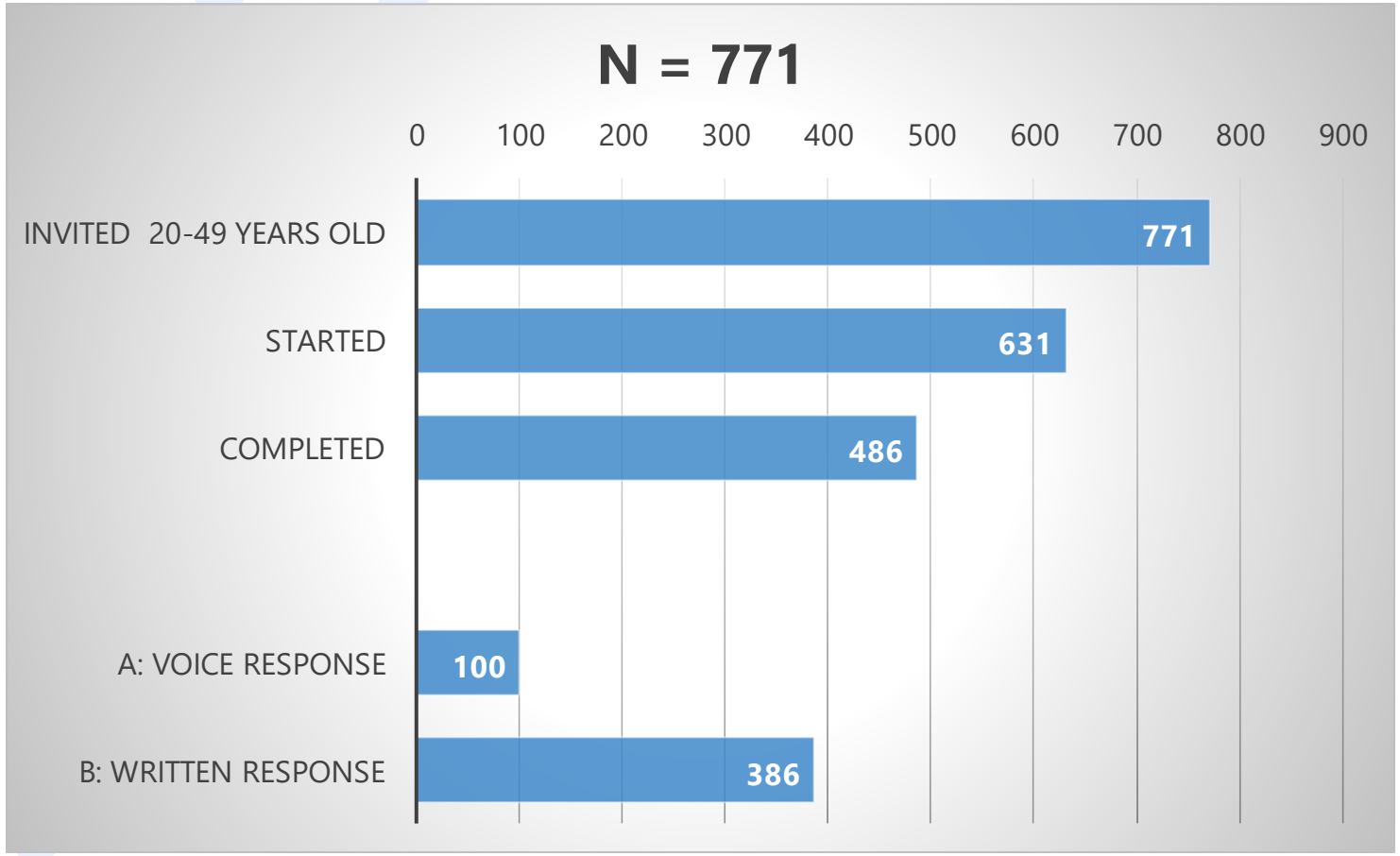
Van den Heuvel et al., 2023

Feasibility approach of CARI in CAWI

- Speech and text input comparison
- Quality of audio and ASR transcriptions
- Sentiment Analysis
- Topic Modelling

Connecting Humanities and Social Sciences: Applying Language and Speech Technology to Online Panel Surveys.

Henk van den Heuvel, Martijn Bentum, Simone Wills, Judith C. Koops, 2023



Fielded in April 2021

SSHOC quasi-exp with 2 conditions

- N = 771 invited
- 20 – 49 years old

Response

- 631 (82%) started
- 486 (63%) completed

Response conditions

- 100 (21%) voice response
- 386 (79%) written response

Collected audio files

- 2379 audio files
- 1796 audio and matched transcription
- 7 hours and 15 minutes of audio

Speech and text input comparison

	Speech	Keyboard
# responses	1,665	4,322
median # words	16	9
average # words	25.96	12.09
max # words	139	209
total # words	43,216	52,249
median # content words	13	6
average # content words	18.9	8.55
total # content words	30,539	36,974
percentage content words	70.69%	70.76%

Table 1: Comparison of speech and keyboard input modality for questionnaire answers.

Respondents provide **longer answers** with **Speech to Text** compared to keyboard input.

Modalities do not appear to influence percentage of content words.

→ **Talk more, but not more actual content?**

Audio & ASR quality

Label	Frequency	Percentage
Good	338	56.90%
Average	187	31.48%
Poor	53	8.92%
Very poor	16	2.69%

Table 2: Perceptual assessment of the audio recordings

Label	WER	subs	del	ins
Questfox	24.7	9.19	13.97	1.54
DC	34.34	14.51	17.12	2.71
OH	36.51	15.54	18.23	2.73
PD	34.26	14.48	17.07	2.71

Table 3: Performance in Word Error Rate (WER) for the various speech recognisers.

Almost 90% of recordings are good or average acoustic quality, well suited for ASR.

Questfox ASR outperforms the other engines by around 10 - 12%.

Even though 90% of recordings are of sufficient quality for ASR, the Word Error Rate is 25%, indicating that there is **ample room** for **improvement** of the **ASR engines**.

(sentiment analyses & topic modelling)

Discussion and take home messages

1. Speech to Text / ASR in online surveys is possible
2. Response bias
   - unwillingness to participate
   - technical inability or issues
3. ASR and audio quality
   - accuracy and validity issues
   - background noise and social context
4. Audio is valuable data source for researchers, but what's in it for respondents? → make it fun and offer an incentive!
5. What other (better?) S2T tools or methods are suitable for online surveys?

Questions?

Joris Mulder – joris.mulder@centerdata.nl


CENTERDATA
research institute