

# Investigating paradata for one of the largest surveys in Sweden

Andreea Bolos

Viktor Dahl

Sofia Holsendahl

Statistics Sweden

SCB

# Content

- **Introduction**
  - Aim & questions
  - Method
- **Results**
  - Contact process
  - Device type
  - Questionnaire navigation
    - > Response time

# Aim

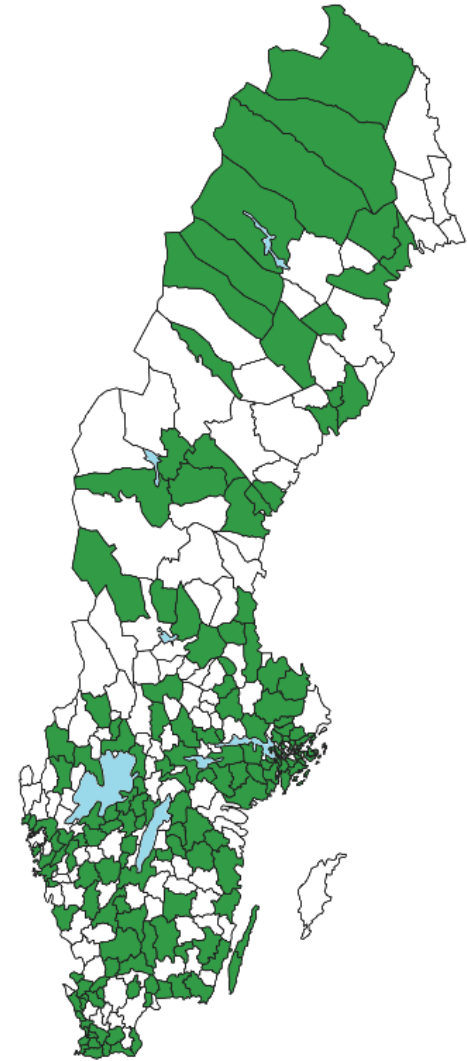
- To study different **typologies of web paradata** and their role in improving data collection efficiency and ensuring a high data quality.

# Questions

- Are specific **socio-demographic** characteristics of the study sample associated with longer or shorter response times?
- Are specific **device types** associated with longer or shorter response times?
- Do the **citizens' attitudes towards their municipality** have any affect on the response time?

# Sample and method

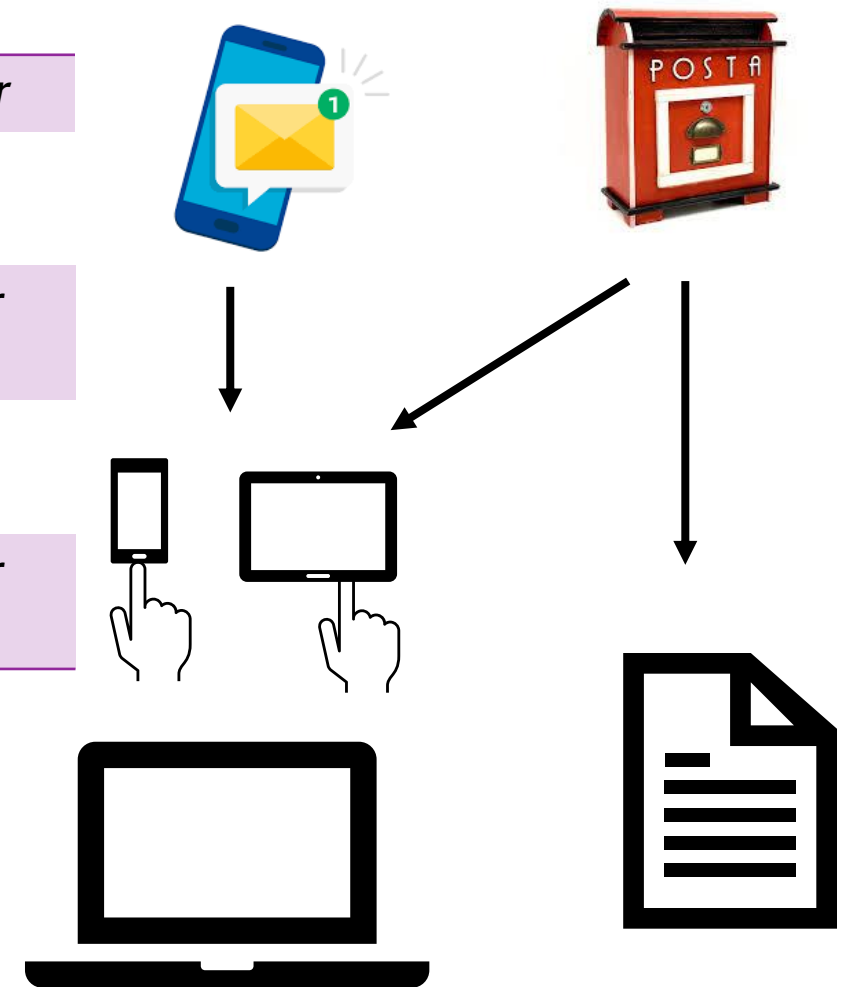
- **Sample**
  - 160 municipalities
    - 18 years-old and older
    - $n = 208\ 000$
    - response rate: 36 percent
      - Expected average answer time 20 min
- **Method**
  - Exploratory
    - > Contact information paradata
    - > Device type paradata
    - > Questionnaire navigation paradata
  - Linear regression model
    - > Time as dependent variable
    - > Demographics, device type and municipality score as independent variables



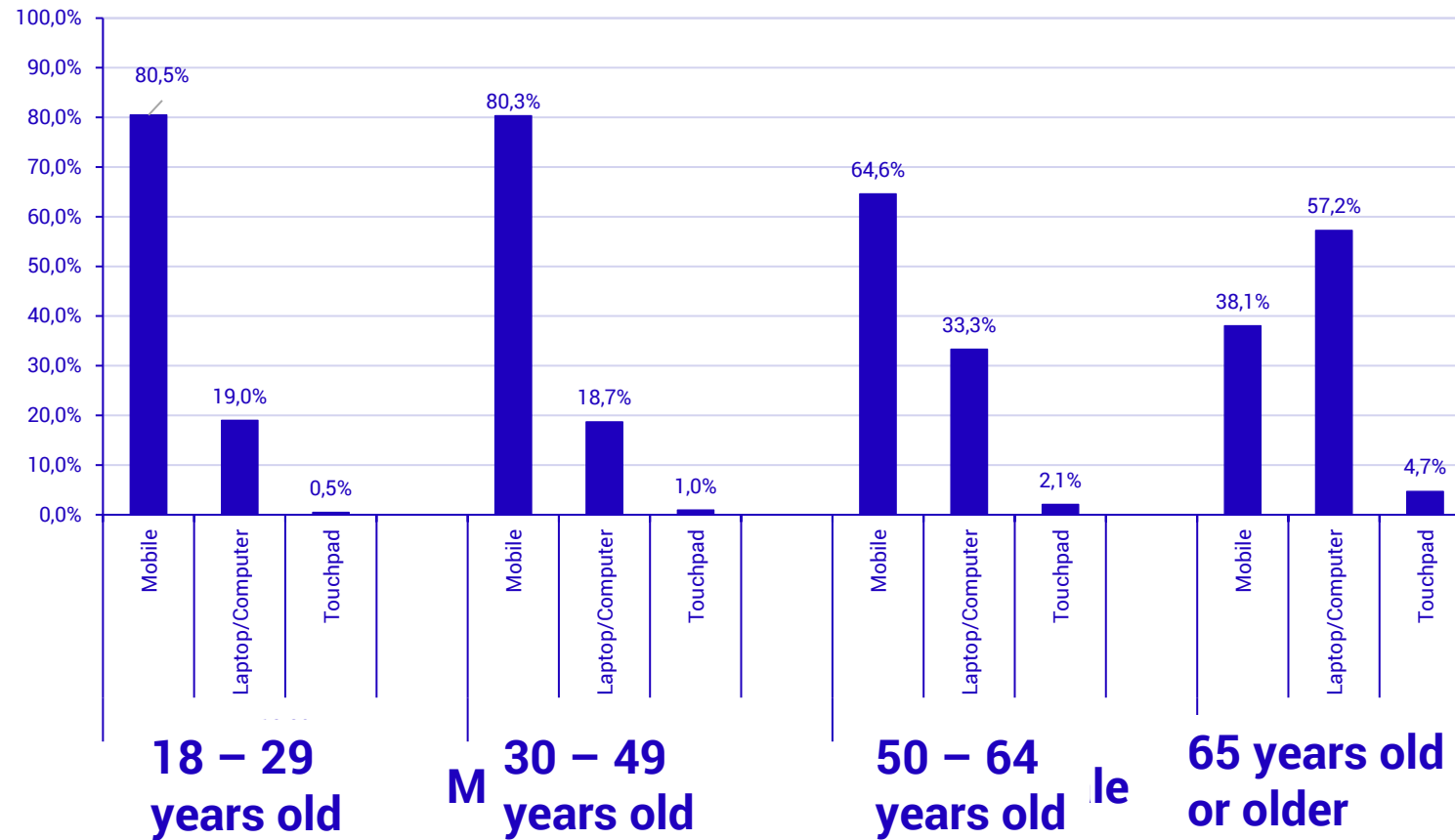
# Contact strategy

**Table 1. Contact attempts**

<i>Send out</i>	<i>Persons under 65 years old</i>	<i>Persons 65 years old or older</i>
1	Invitation with login to the web survey	Invitation with login to the web survey
2	Reminder	Reminder including a paper questionnaire
3	Reminder including a paper questionnaire	Reminder
4	Reminder	Reminder including a paper questionnaire



# Results 1/5 – Device type



# Results 2/5 – Active time per gender and age

	N	Mean	Std Deviation	Median	Variance
<b>Gender</b>					
• Male	14 316	20,7	11,1	17,8	124,0
• Female	15 862	20,1	11,1	17,1	122,5
<b>Age</b>					
• 18–29	4 131	19,1	11,6	15,7	133,6
• 30–49	8 141	19,2	12,0	15,9	143,5
• 50–64	11 078	19,9	10,3	17,2	105,9
• 64 or older	6 828	23,4	10,4	20,9	109,1
<b>Total</b>	<b>30 178</b>	<b>20,4</b>	<b>11,1</b>	<b>14,6</b>	<b>123,3</b>

## Results 3/5 – Active time per device

	<b>N</b>	<b>Mean</b>	<b>Std Deviation</b>	<b>Median</b>	<b>Variance</b>
<b>Laptop/computer</b>	9848	22,3	12,1	19,2	146,2
<b>Mobile</b>	19 501	19,3	10,4	16,6	107,5
<b>Touch Pad</b>	645	21,8	10,2	17,8	103,1
<b>Total</b>	<b>29 994</b>	<b>20,3</b>	<b>11,1</b>	<b>17,43</b>	<b>122,1</b>



*Note.* Sample - only for those who did not change device



# Results 4/5 – Outliers

	Less than 5 min n (%)	5<=active time<=60 n (%)	Above 60 min n (%)
<b>Age</b>			
Low	15 (57,7)	4 698 (13,4)	152 (13,6)
Medium	19 (37,9)	14 988 (20,9)	173 (53,8)
High	3 (24,3)	10 282 (34,8)	94 (27,4)
<b>Education</b>	2 (6,9)	6 746 (22,6)	80 (23,6)
<b>Country of birth</b>	7 (24,1)	3 231 (10,8)	70 (20,7)
Sweden	19 (67,9)	18 326 (48,3)	202 (68,4)
Outside Sweden	10 (37,5)	13 493 (41,7)	137 (40,4)
<b>N</b>	<b>29</b>	<b>29 810</b>	<b>339</b>

# Results 5/5 Linear regression model

Parameter	Estimate	Pr> t
<b>Intercept</b>	25,45	***
<b>Gender</b>		
• Male	0,32	ns
• Female	Ref	Ref
<b>Age</b>		
• 18–29	-4,71	***
• 30–49	-3,47	***
• 50–64	-2,63	ns
• 65 or older	Ref	Ref
<b>Country of birth</b>		
• Sweden	-5,33	***
• Outside Sweden	Ref	Ref
<b>Education</b>		
• Primary	1,11	***
• Secondary	0,64	***
• Higher	Ref	Ref
<b>Income</b>		
• Low	2,72	***
• Medium	1,85	***
• High	Ref	Ref
<b>Device</b>		
• Laptop	1,95	***
• Mobile	Ref	Ref
<b>Municipality score</b>	-0,12	ns

Model: p <0,0001

NB: Comparisons significant at the 0.05 level are indicated by \*\*\*



# Conclusions and further work

- Using response time to clean data
  - Nature of responses
    - > Fast responses = attentive or not paying close attention?
    - > Slow responses = attentive, distracted or confused?
  - Additional independent factors
    - > Item/partial nonresponse (high vs. low)
    - > Time of day (morning, day, evening, night)
- Ways forward
  - Compare drop-outs with "send-ins"
  - Address other forms of paradata
    - > Item/partial nonresponse
    - > Answer time per question



Andreea Bolos  
Statistics Sweden  
[andreea.bolos@scb.se](mailto:andreea.bolos@scb.se)

Viktor Dahl  
Statistics Sweden  
[viktor.dahl@scb.se](mailto:viktor.dahl@scb.se)

Sofia Holsendahl  
Statistics Sweden  
[sofia.holsendahl@scb.se](mailto:sofia.holsendahl@scb.se)