



Centerdata
Institute for data collection and research



Outline

1. Online interviewing for scientific research
2. Can we do better?
 - Is it feasible to cover the general population with an Internet panel?
 - What about data quality?
 - Is it useful for longitudinal research?
3. Data quality:
 - Validity and reliability (MTMM)
 - In comparison with CATI and CAPI

Experience of two years

LISS panel

Longitudinal Internet studies in the Social
Sciences



Centerdata
Institute for data collection and research



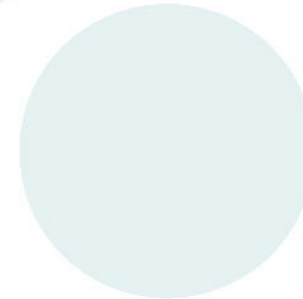
LISS panel

- Online panel of 5000 households
- For scientific use, at no cost
- Questionnaires every month
- Seven year longitudinal study
- Probability sample including households without internet access



Disadvantages online surveys

- Noncoverage:
 - Persons without internet are usually excluded
- No probability sample:
 - Visitors of specific websites
 - Selfselection (interested, strong opinions, specific goal)



Solution:

- Use internet for interviews, not for selection
- Draw from population frame
- Invite only selected respondents



How? Example LISS panel

- Probability sample drawn from address sampling frame of SN
- Includes household without internet access (less than 15%): Equipment
- Contacted by letter, telephone or visit



Online interviewing and data quality

1. MTMM study of CAWI, CATI, CAPI data, using the same sample
2. Focus not on sampling and selectivity, but on effect of mode itself



“Real” mode effect

- Self-administered
 - Socially desirable answering
 - Help, explanations
- Question formats
 - Example: DK not read aloud
 - Shift to visual design
- Technical developments
 - Screen, browser, programs (flash)
- Human-computer interaction
 - Cognitive and functional ability respondent
 - » Example: mousing efficiency segmentation



Comparison Cati, Capi, Cawi

NOT:

- Coverage, response, selection bias
- Answer distributions, means

BUT:

- Validity and reliability of data =
- Mode specific random and systematic error variance



Multitrait multimethod model

11

11-6-2009



The true-score mtmm model

$$\mathbf{y}_i = \mathbf{h}_i \mathbf{T}_i + \mathbf{e}_i \quad (1)$$

$$\mathbf{T}_i = \mathbf{b}_{ij} \mathbf{F}_j + \mathbf{g}_{ik} \mathbf{M}_k \quad (2)$$



The true-score mtmm model

Reliability

$$y_i = h_i T_i + e_i \quad (1)$$

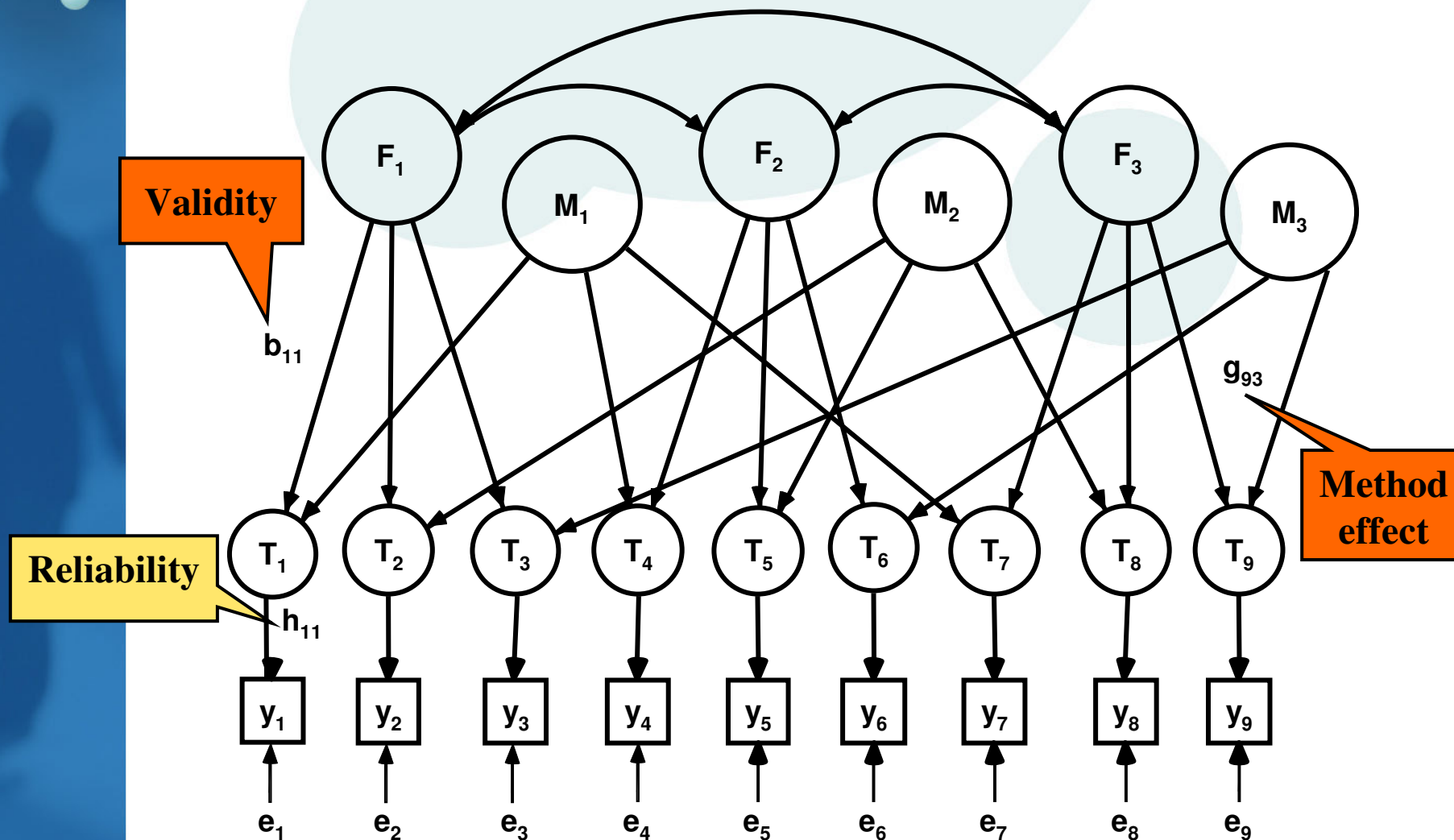
$$T_i = b_{ij} F_j + g_{ik} M_k \quad (2)$$

Validity

Method
effect

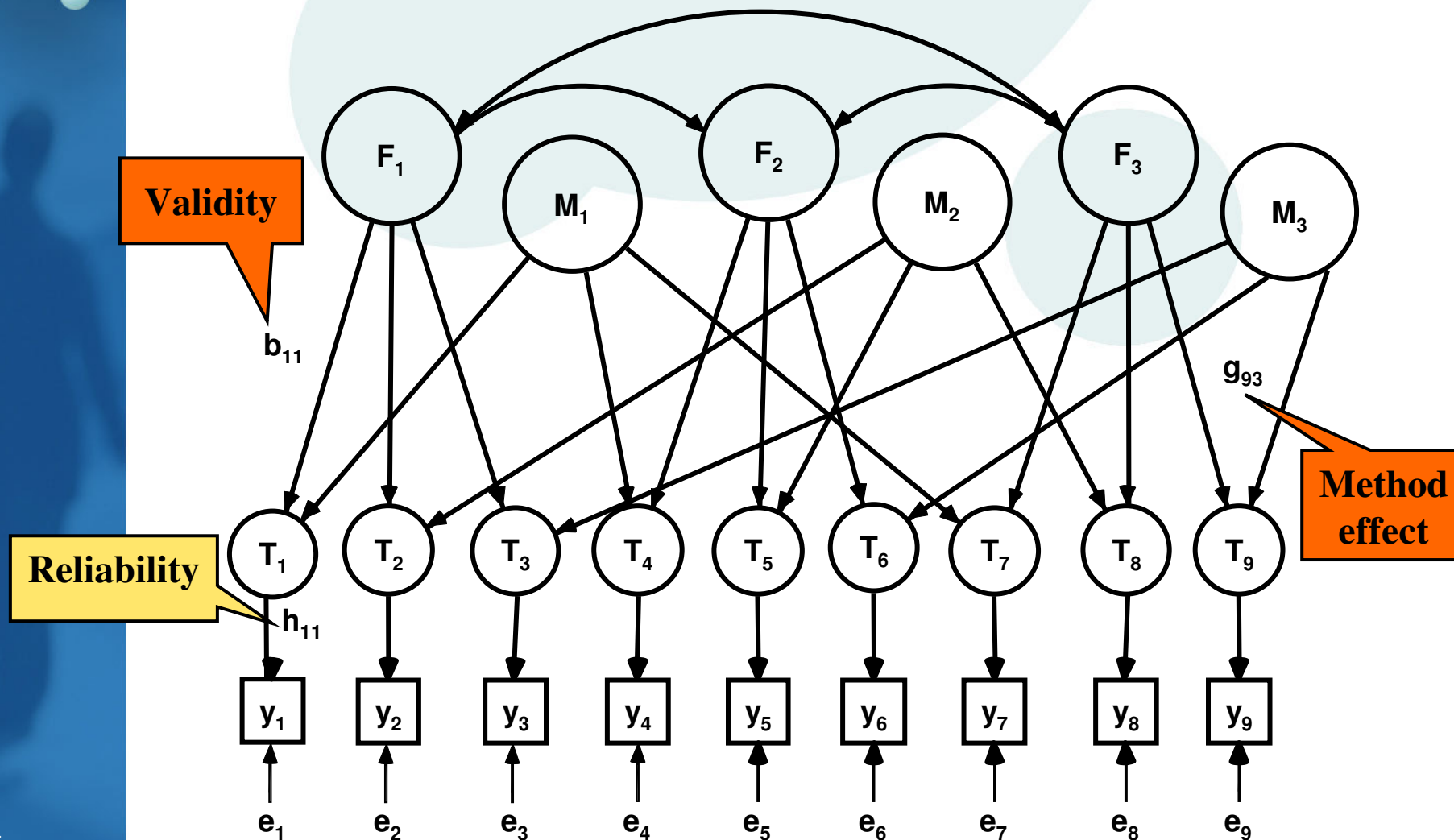


The true-score mtmm model





Multiple traits Multiple methods





Multiple traits, Multiple methods



Example: Societal Participation

Measure	Trait	Method
1	Voluntary work	Cati
2	Active member of organisations	Cati
3	Hours sport per week	Cati
	Some time later	
4	Voluntary work	Cawi
5	Active member of organisations	Cawi
6	Hours sport per week	Cawi

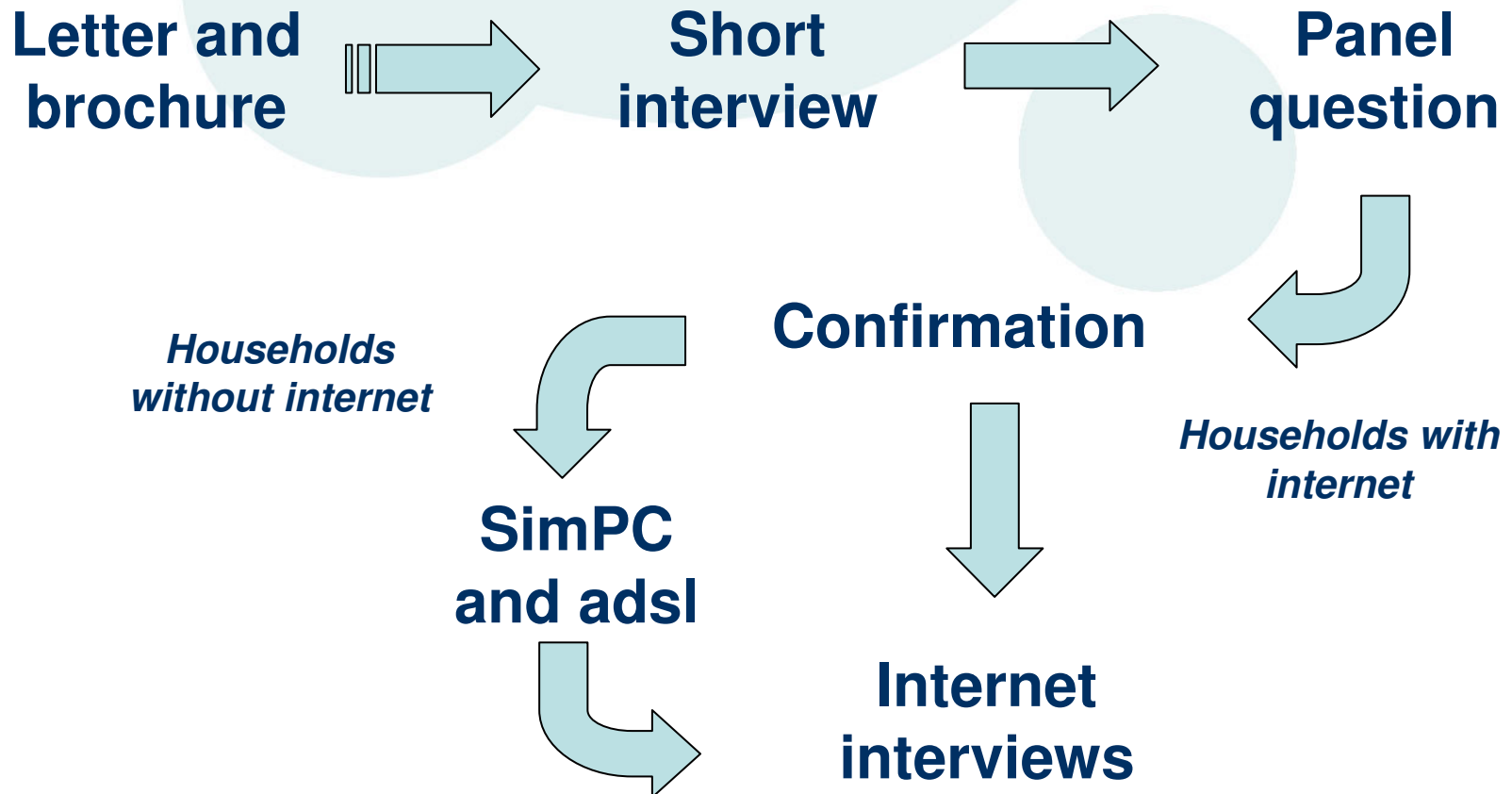


Example LISS panel

- Probability sample drawn from address sampling frame of SN
- Includes household without internet access (less than 15%): Equipment
- Contacted by letter, telephone or visit



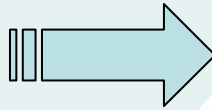
Recruitment of LISS panel





Recruitment

Letter and
brochure



Short
interview



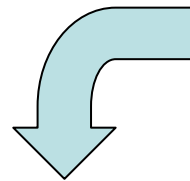
Controlgroup
question

Cati if telnr

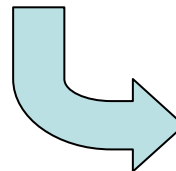
Capi if not

Confirmation

*Households
without internet*



SimPC
and adsl



Internet
interviews

*Households with
internet*





Experimental design

	2007 Feb – Nov	2007 Nov	Population	Net N
Group 1	Cati	Cawi	Telnr	2038
Group 2	Capi	Cawi	Telnr + no telnr	111 + 642



Experimental design

	2007 Feb – Nov	2007 Nov	Population	Net N
Group 1	Cati	Cawi	Telnr	2038
Group 2	Capi	Cawi	Telnr + no telnr	111 + 642



Experimental design

	2007 Feb – Nov	2007 Nov	Population	Net N
Group 1	Cati	Cawi	Telnr	2038
Group 2	Capi	Cawi	Telnr + no telnr	111 + 642



Experimental design

	2007 Feb – Nov	2007 Nov	Population	Net N
Group 1	Cati	Cawi	Telnr	2038
Group 2	Capi	Cawi	Telnr + no telnr	111 + 642



Split ballot multitrait multimethod model

25

11-6-2009



Results

- **9 traits x 3 methods = 27 measures**
- **27 validity coefficient**
- **27 method effects**
- **27 reliability coefficients**

Secondary analysis



Secondary analysis of mtmm estimates

Factor	N	Validity coefficient	
		Estimated marginal mean	Eta sqd
Interview mode			.26*
Cati	9	.94	
Capi	9	.91	
Cawi	9	.93	
Topic			.87*
Interest in news and politics	9	.98	
Personality	6	.90	
Life satisfaction	3	.79	
Societal participation	6	.98	
Sport participation	3	.99	
Grand mean (sd)		.93 (.01)	
Adjusted R Squared			.84

* $p < 0.05$



Secondary analysis of mtmm estimates

Number of coefficients!

Factor	N	Validity coefficient	
		Estimated marginal mean	Eta sqd
Interview mode			.26*
Cati	9	.94	
Capi	9	.91	
Cawi	9	.93	
Topic			.87*
Interest in news and politics	9	.98	
Personality	6	.90	
Life satisfaction	3	.79	
Societal participation	6	.98	
Sport participation	3	.99	
Grand mean (sd)		.93 (.01)	
Adjusted R Squared			.84

* $p < 0.05$



Secondary analysis of mtmm estimates

Factor	N	Validity coefficient	
		Estimated marginal mean	Eta sqd
Interview mode			.26*
Cati	9	.94	
Capi	9	.91	
Cawi	9	.93	
Topic			.87*
Interest in news and politics	9	.98	
Personality	6	.90	
Life satisfaction	3	.79	
Societal participation	6	.98	
Sport participation	3	.99	
Grand mean (sd)		.93 (.01)	
Adjusted R Squared			.84

* $p < 0.05$



Secondary analysis of mtmm estimates

Factor	N	Validity coefficient	
		Estimated marginal mean	Eta sqd
Interview mode			.26*
Cati	9	.94	
Capi	9	.91	
Cawi	9	.93	
Topic			.87*
Interest in news and politics	9	.98	
Personality	6	.90	
Life satisfaction	3	.79	
Societal participation	6	.98	
Sport participation	3	.99	
Grand mean (sd)		.93 (.01)	
Adjusted R Squared			.84

* $p < 0.05$



Secondary analysis of mtmm estimates

Factor	N	Validity coefficient	
		Estimated marginal mean	Eta sqd
Interview mode			.26*
Cati	9	.94	
Capi	9	.91	
Cawi	9	.93	
Topic			.87*
Interest in news and politics	9	.98	
Personality	6	.90	
Life satisfaction	3	.79	
Societal participation	6	.98	
Sport participation	3	.99	
Grand mean (sd)		.93 (.01)	
Adjusted R Squared			.84

* $p < 0.05$



Secondary analysis of mtmm estimates

Factor	N	Reliability coefficient	
		Estimated marginal mean	Eta sqd
Interview mode			.59*
Cati	9	.76	
Capi	9	.67	
Cawi	9	.82	
Topic			.40*
Interest in news and politics	9	.74	
Personality	6	.78	
Life satisfaction	3	.66	
Societal participation	6	.79	
Sport participation	3	.79	
Grand mean (sd)		.75 (.01)	
Adjusted R Squared			.58

* $p < 0.05$



Secondary analysis of mtmm estimates

Factor	N	Reliability coefficient	
		Estimated marginal mean	Eta sqd
Interview mode			.59*
Cati	9	.76	
Capi	9	.67	
Cawi	9	.82	
Topic			.40*
Interest in news and politics	9	.74	
Personality	6	.78	
Life satisfaction	3	.66	
Societal participation	6	.79	
Sport participation	3	.79	
Grand mean (sd)		.75 (.01)	
Adjusted R Squared			.58

* $p < 0.05$



Secondary analysis of mtmm estimates

Factor	N	Reliability coefficient	
		Estimated marginal mean	Eta sqd
Interview mode			.59*
Cati	9	.76	
Capi	9	.67	
Cawi	9	.82	
Topic			.40*
Interest in news and politics	9	.74	
Personality	6	.78	
Life satisfaction	3	.66	
Societal participation	6	.79	
Sport participation	3	.79	
Grand mean (sd)		.75 (.01)	
Adjusted R Squared			.58

* $p < 0.05$



Conclusions

- 😊 **Panel data collected using a web interview are at least as valid and reliable as data collected using more traditional modes of interviewing, when the sample is kept constant**
- ☹️ **Error variance due to data collection techniques cannot be ignored**



Conclusions

Scientific research: Can we do better?

- ✓ Is it feasible to cover the general population with an Internet panel?
- ✓ [What about data quality?](#)
- ✓ Is it useful for longitudinal research?

Yes we can!





LISS Core Study

- Household and family
- Economic situation and housing
- Work and schooling
- Social integration and leisure
- Health
- Personality
- Religion and ethnicity
- Politics and values

LONGITUDINAL: repeated each year

LONG: spread over 8 months!



Use of LISS panel and data

- Available for academic use
 - Send in your own proposal (now about 40)
 - Use available data (now about 15 datasets)
- Disseminated through website

www.lissdata.nl

- Read more



Centerdata
Institute for data collection and research



Design control analyses

	2007 Feb – Nov	2007 Nov	Population	Net N
Group 1	Cati	Cawi	Telnr	2038
Group 2	Capi	Cawi	Telnr + no telnr	111 + 642



Design control analyses

		Eta sqd	
Factor	N	Validity coefficient	Reliability coefficient
Time between interviews		.03	.02
1 to 8 months	54		
Less than 2 months	18		
Subpopulation		.03	.01
All	45		
Known landline nr	18		
Mode * Time		.02	.02
Mode * Subpopulation		.04	.02

* $p < 0.05$