

Psychometric results from the Wellbeing Multi- Instrument Comparison (W-MIC) study

Agenda

Background and project aim

Methodology

Results

Discussion

Q&A

Erasmus

Erasmus School of
Health Policy
& Management

Background and research aim

Erasmus University Rotterdam



Background (1/2)



Economic evaluations (EE) advocated to be used in long-term care (LTC) and non-curative settings



Current outcome measures are inadequate given their focus on (physical) health



Extending applications of EE to LTC settings will require focusing on broader outcomes, such as well-being

Erasmus

Background (2/2)



Several well-being questionnaires have been developed to use in non-curative care to evaluate technology and services



These questionnaires are at different stages of development and have never been compared altogether in the same context/population



It is still unclear which instruments should be used to assess the effectiveness of an intervention targeting well-being in a specific context



This question led to the project named "The Well-being Multi-Instrument Comparison" (W-MIC) study

Erasmus

Erasmus School of
Health Policy
& Management

Methodology

Erasmus University Rotterdam



Methodology



Study design

Cross-sectional survey
with follow-up
measurement



Population

7000 US and NL
participants, with quotas
for specific subgroups
(general, older, targeted
health conditions)



Measures

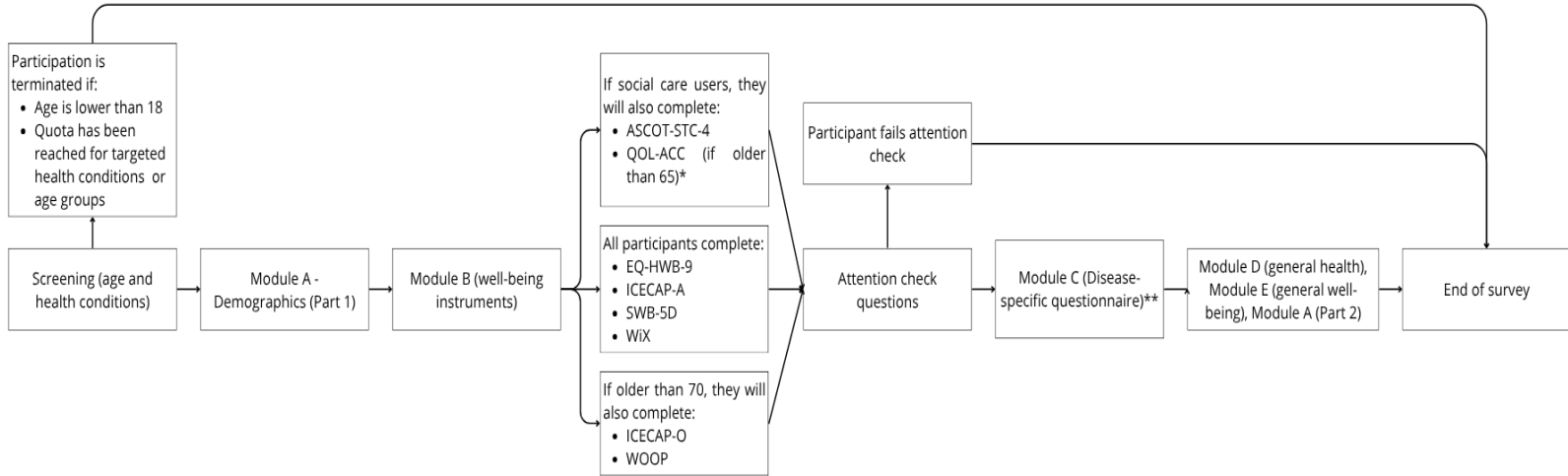
Demographics (Module A), Well-
being instruments (Module B),
Clinical severity (Module C),
General health (Module D),
General well-being (Module E)



Data collection

Dynata hosted and
recruited participants

Survey logic



Statistical analysis

- Ceiling/floor effects
 - Convergent validity
 - Known-groups validity
 - Test-retest reliability
-
- To facilitate comparison between instruments with different scoring systems, level sum scores were linearly transformed to a standardized 0–100 scale.

Erasmus

Erasmus School of
Health Policy
& Management

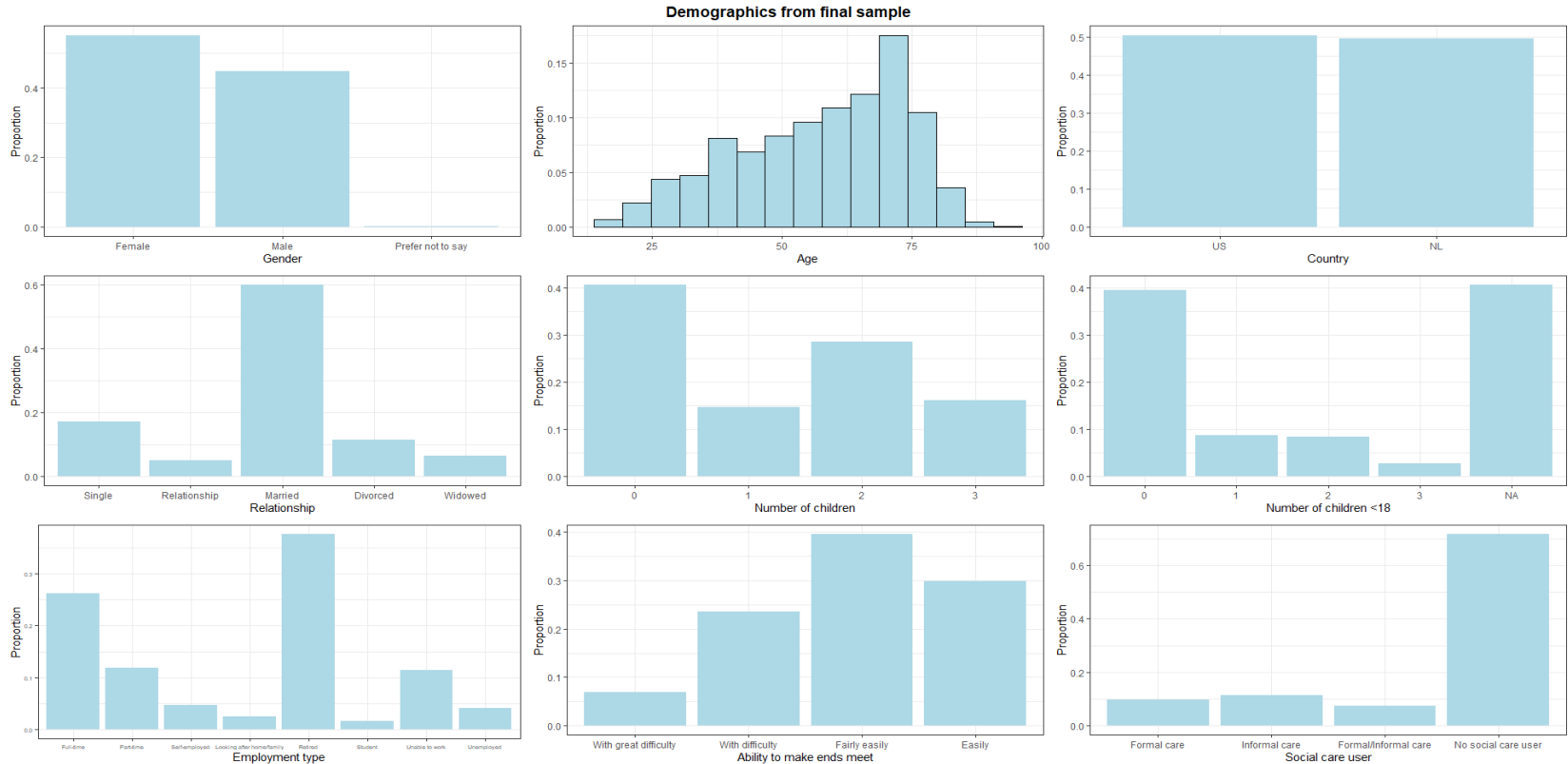
Results

Erasmus University Rotterdam

The Erasmus University logo, featuring the word "Erasmus" in a white, cursive script font.

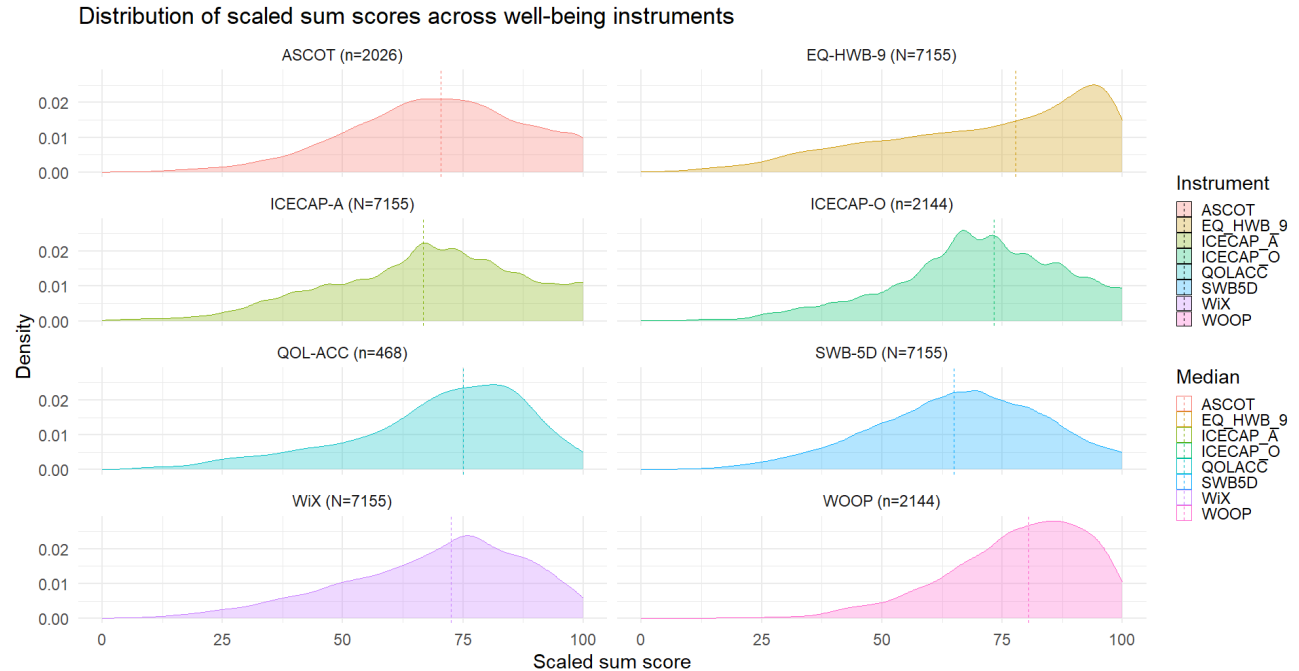
Demographics

- 7155 final sample (US: n= 3606, 50%; NL: n=3549, 50%)



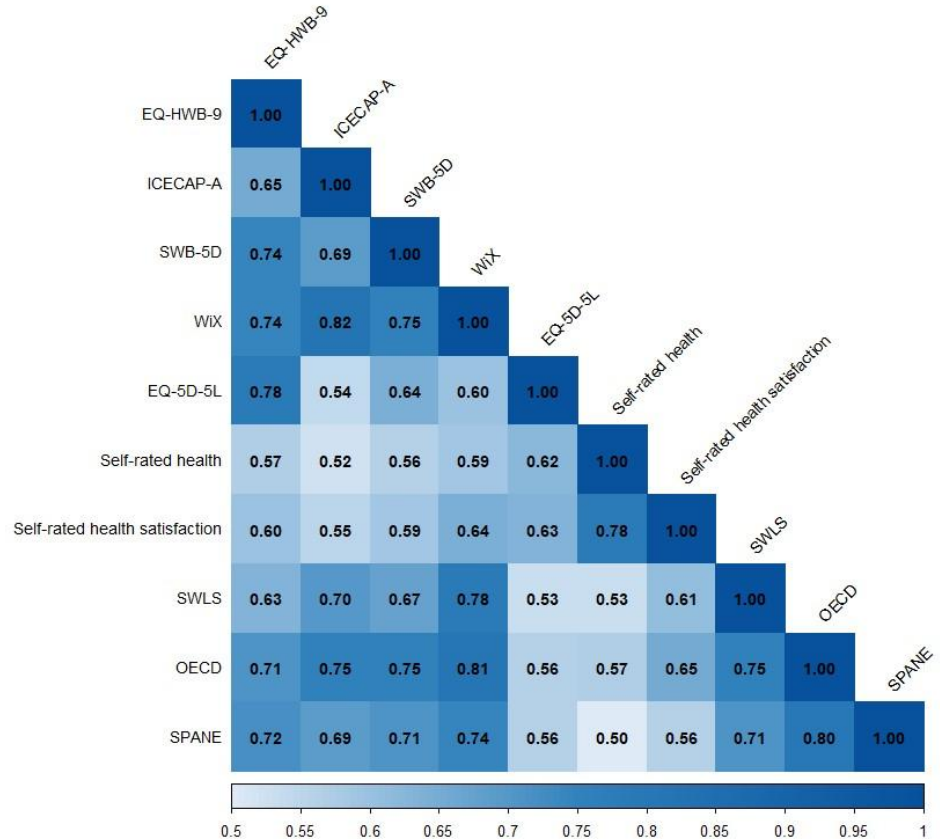
Descriptive statistics

- Participants had moderate to high levels of well-being across instruments
- More negatively skewed in the EQ-HWB-9 and WOOP



Convergent validity

- All well-being instruments showed strong correlations among them
- EQ-HWB-9 was more strongly correlated with EQ-5D-5L
- WiX was more strongly correlated with ICECAP-A and general well-being questionnaires



Known-group validity

- All well-being instruments were able to distinguish between clinical groups
- EQ-HWB-9 and SWB-5D did not detect differences among social care users in post-hoc comparison using Bonferroni correction

Group	EQ-HWB-9	ICECAP-A	SWB-5D	WiX
Depression				
Clinical (n=634) – Mean (SD)	45.01 (15.26)	48.23 (17.84)	47.71 (14.24)	47.97 (16.52)
Non-clinical (n=412) – Mean (SD)	72.26 (14.71)	66.31 (17.30)	65.36 (14.03)	67.28 (14.23)
Mann-Whitney U results – U (p) [r]	27533 (<0.001) [0.67]	60313 (<0.001) [0.46]	49571.5 (<0.001) [0.53]	47054 (<0.001) [0.54]
Arthritis				
High severity (n=579) – Mean (SD)	55.72 (19.17)	59.56 (19.39)	56.60 (14.14)	61.03 (18.14)
Moderate severity-Near remission (n=499) – Mean (SD)	81.17 (14.10)	75.55 (15.15)	72.89 (13.68)	77.05 (13.60)
Mann-Whitney U results – U (p) [r]	40786.5 (<0.001) [0.62]	74783.5 (<0.001) [0.42]	59137 (<0.001) [0.51]	67287 (<0.001) [0.46]
Sleep				
High sleep disturbance (n=505) – Mean (SD)	56.89 (20.33)	56.92 (19.91)	56.78 (17.11)	56.68 (18.88)
Low sleep disturbance (n=571) – Mean (SD)	75.58 (19.02)	70.99 (19.12)	68.99 (16.72)	71.35 (17.04)
Mann-Whitney U results – U (p) [r]	71801.5 (<0.001) [0.43]	87189.5 (<0.001) [0.34]	87060.5 (<0.001) [0.34]	79889 (<0.001) [0.39]
Skin				
Moderate to Extreme effect on life (n=408) – Mean (SD)	60.59 (21.61)	63.94 (20.29)	61.99 (16.69)	62.34 (19.52)
Small-no effect on life (n=673) – Mean (SD)	77.96 (17.35)	70.69 (17.93)	70.25 (16.11)	71.14 (16.07)
Mann-Whitney U results – U (p) [r]	72485 (<0.001) [0.40]	110711.5 (<0.001) [0.16]	97203.5 (<0.001) [0.25]	100264.5 (<0.001) [0.23]
Heart				
High physical impact (n=561) – Mean (SD)	61.51 (19.59)	61.16 (18.98)	57.09 (15.30)	61.88 (17.85)
Low physical impact (n=547) – Mean (SD)	88.27 (11.04)	78.99 (13.91)	75.75 (12.33)	80.54 (12.11)
Mann-Whitney U results – U (p) [r]	31012.5 (<0.001) [0.69]	70275 (<0.001) [0.47]	51867.5 (<0.001) [0.58]	57110 (<0.001) [0.54]
General Health				
Poor health (n=4501) – Mean (SD)	63.74 (21.32)	61.02 (19.45)	59.39 (16.41)	61.74 (18.41)
Good health (n=2654) – Mean (SD)	86.02 (15.20)	79.33 (14.95)	77.22 (13.77)	81.15 (12.36)
Mann-Whitney U results – U (p) [r]	9757852 (<0.001) [0.53]	9182292 (<0.001) [0.45]	9521704 (<0.001) [0.50]	9693975 (<0.001) [0.52]
Receiving care				
Formal care users (n=700)	57.37 (22.99)	63.27 (22.58)	58.01 (17.30)	63.88 (21.01)
Formal and informal care users (n=519)	57.65 (21.51)	61.93 (20.99)	58.90 (16.65)	62.42 (19.96)
Informal care users (n=807)	57.24 (19.97)	59.45 (18.80)	57.41 (16.32)	60.67 (18.29)
Non care recipient (n=5129)	77.78 (19.51)	70.35 (19.07)	69.16 (17.08)	71.59 (17.90)
Kruskal-Wallis results – H (p) [η ²]	1224.045 (<0.001) [0.17]	318.5162 (<0.001) [0.04]	612.6007 (<0.001) [0.09]	375.3747 (<0.001) [0.05]

Effect size interpretation: small ($0.01 \leq \eta^2 < 0.06$, $0.1 \leq r < 0.24$); moderate ($0.06 \leq \eta^2 < 0.14$, $0.24 \leq r < 0.37$); strong ($\eta^2 \geq 0.14$, $r \geq 0.37$)

Test-retest reliability

- All well-being instruments showed reliable assessment over time (ICC>0.70)

Instrument	Correlations main and retest sample	Complete agreement	ICC (95% CI)
No change in health (n=790)			
<i>EQ-HWB-9</i>	0.84	20.00%	0.84 (0.82-0.86)
<i>ICECAP-A</i>	0.74	24.94%	0.75 (0.72-0.78)
<i>SWB-5D</i>	0.74	21.52%	0.76 (0.73-0.79)
<i>WiX</i>	0.83	13.80%	0.83 (0.80-0.85)
No change in well-being (n=752)			
<i>EQ-HWB-9</i>	0.83	20.61%	0.83 (0.80-0.85)
<i>ICECAP-A</i>	0.72	25.13%	0.74 (0.71-0.77)
<i>SWB-5D</i>	0.73	21.81%	0.75 (0.72-0.78)
<i>WiX</i>	0.82	13.96%	0.83 (0.80-0.85)

Erasmus School of
Health Policy
& Management

Discussion

Erasmus University Rotterdam



Main findings



All instruments demonstrated good psychometric performance (particularly the WiX and EQ-HWB-9)



EQ-HWB-9 and SWB-5D seemed more sensitive to variations in health



ICECAP-A and WiX had higher sensitivity in broader well-being domains

Erasmus

Where would these differences come from?

- They may be a consequence of the conceptual foundation of these instruments

EQ-HWB

- Based on Wilson and Cleary's model of health-related quality of life
- Items generated following qualitative review focused on health conditions (5)

SWB-5D

- Originated from Delphi study with overarching theme: "most important dimensions of health-related quality of life" (6)

ICECAP-A

- Based on capability approach
- Items generated in interviews exploring what is important to individuals in their life (7)

WiX

- Items identified with scoping review of existing well-being instruments
- Selected based on theoretical framework of eleven broad domains of well-being (8)

5) Brazier J, Peasgood T, Mukuria C, Marten O, Kreimeier S, Luo N, et al. The EQ-HWB: Overview of the Development of a Measure of Health and Wellbeing and Key Results. *Value Health*. 2022 Apr 1;25(4):482–91.

6) Pietersma S, De Vries M, Van Den Akker-van Marle ME. Domains of quality of life: results of a three-stage Delphi consensus procedure among patients, family of patients, clinicians, scientists and the general public. *Qual Life Res*

7) Al-Janabi H, Flynn TN, Coast J. Development of a self-report measure of capability wellbeing for adults: the ICECAP-A. *Qual Life Res Int J Qual Life Asp Treat Care Rehabil*. 2012 Feb;21(1):167–76.

8) Voormolen DC, Bom JAM, de Bekker-Grob EW, Brouwer WBF, van Exel J. Development and Content Validation of the 10-item Well-being Instrument (WiX) for use in Economic Evaluation Studies. *Appl Res Qual Life*. 2024 Apr 1;19(2):381–413.



Erasmus School of
Health Policy
& Management

Thank you!

Erasmus University Rotterdam

The Erasmus University logo, featuring the word "Erasmus" in a white, elegant, cursive script font.

Appendix

Conceptual Coverage

Instrument	Mental health	Physical health	Cognitive health	Relationships	Support	Living environment	Safety	Finance	Relaxation and leisure	Activities	Acceptance and resilience	Independence	Self-worth	Personal hygiene	Life enjoyment / satisfaction	Quality of care	Thinking of future
EQ-HWB	x	x	x	x						x		x					
WiX	x	x		x	~	x	x	x	x	x	~	x	x				
WOOP	x	x	~	x	x	x		x		~	x	x	x				
SWB-5D	x	~		x						~	~	x					
ASCOT				x		x	x		~	x		x	x	x		~	
QOL-ACC	x	x		x					x			x					
ICECAP-A				x	x		x					x	~		x		
ICECAP-O				x					x	~		x	x				x

x indicates full coverage

~ indicates partial coverage



Research aims

1. To concurrently investigate the psychometric properties of multiple well-being instruments

2. To conduct the psychometric investigation separately for different target populations, including general, older, social care and older social care user populations

In-survey quality checks

- Attention check questions
- Dynata's quality check, including:
 - reCAPTCHAs,
 - Geo-IP tracking,
 - Address verification,
 - Digital fingerprinting,
 - IP deduplication
- Dynata's QualityScore:
 - <20 automatically excluded
 - 21-50 manually checked by Dynata

The logo for Erasmus, featuring the word "Erasmus" in a stylized, cursive script font.

Post-hoc data quality checks

- Post-hoc data quality checks:
 - Criteria 1: Inconsistent pain reporting: EQ-HWB-9 (Q4) vs. EQ-5D-5L (Q83)
 - Criteria 2: Inconsistent daily activity reporting: EQ-HWB-9 (Q3) vs. EQ-5D-5L (Q82)
 - Criteria 3: Inconsistent financial situation: WiX (Q10) vs. ability to make ends meet (Q103)
 - Criteria 4: Child count inconsistency: More children <18 (Q99) than total children (Q98)
 - Criteria 5: Completion time <5 minutes
 - Criteria 6: Survey quality score <90
 - Criteria 7: >5 self-reported health conditions
- Participants meeting ≥ 3 criteria were removed
- Participants excluded if provided extreme inconsistencies in Criteria 1 and 2 (>2 Likert points) or in Criteria 3 (e.g., very satisfied but much difficulty)

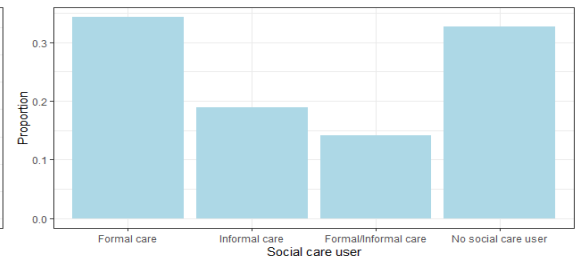
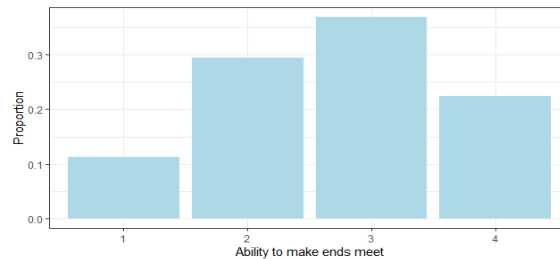
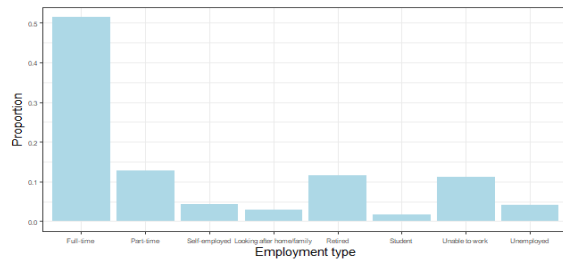
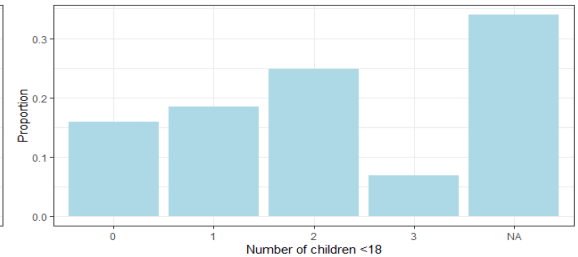
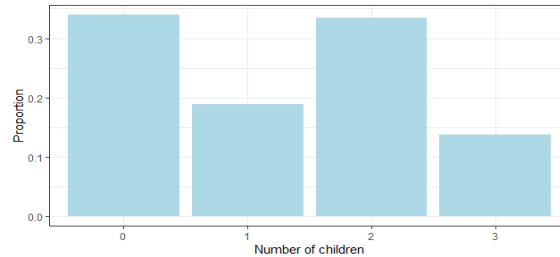
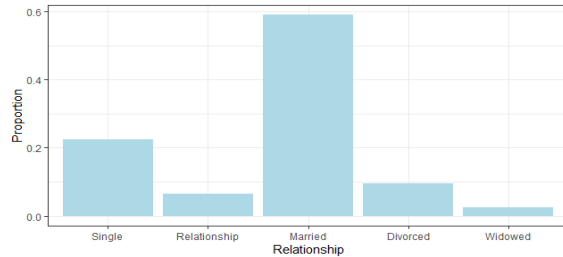
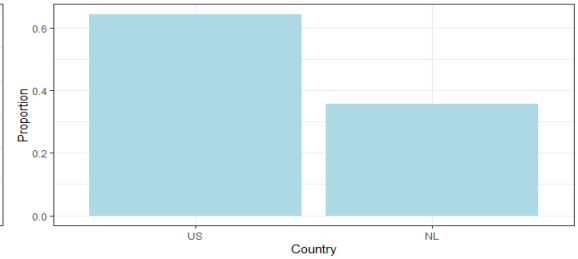
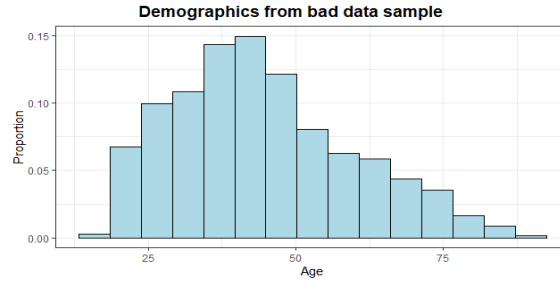
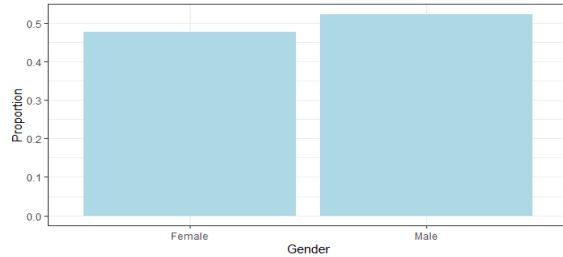
The Erasmus logo is a stylized, handwritten-style signature of the word "Erasmus" in a dark brown or black color, located in the bottom right corner of the slide.

Quality checks

- 7838 participants completed the survey (US: n=4045, 52%; NL: n=3793, 48%)
- 5336 (68%) failed at least one quality check criteria (out of 7)
- 683 (9%) were excluded
 - 560 failed ≥ 3 criteria
 - 202 provided extremely different answers

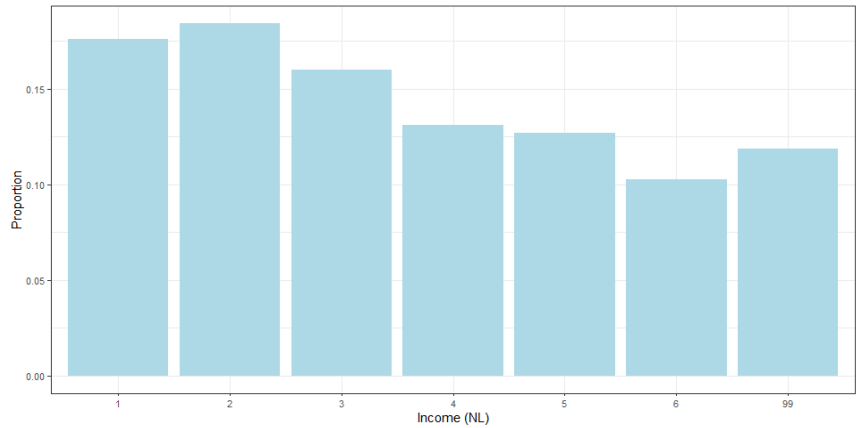
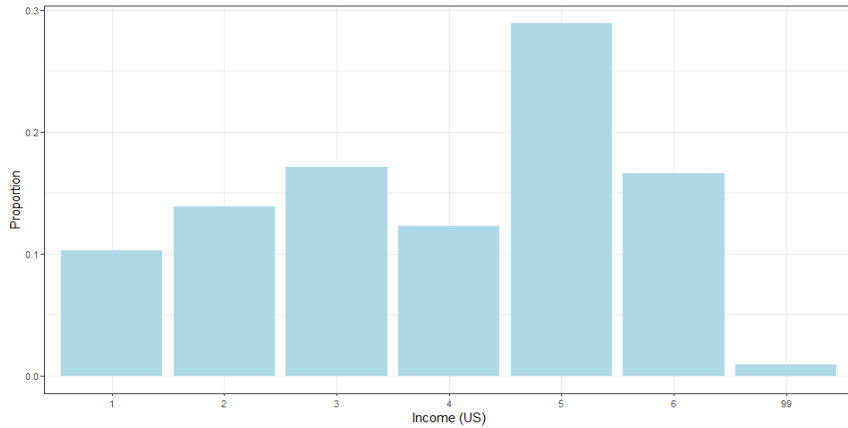
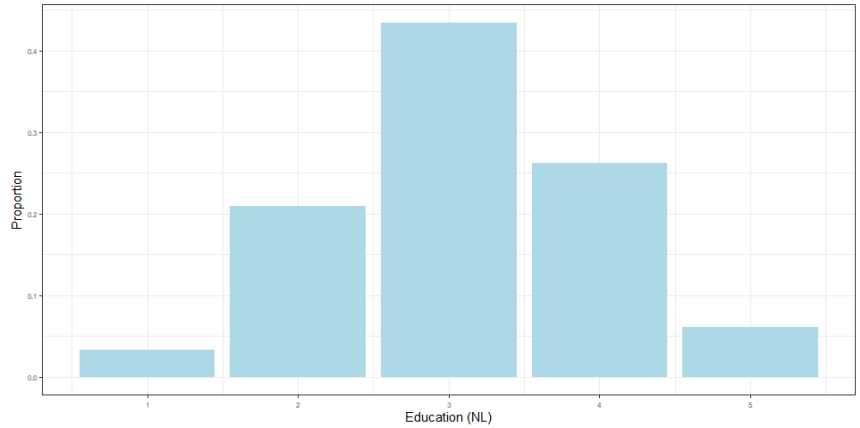
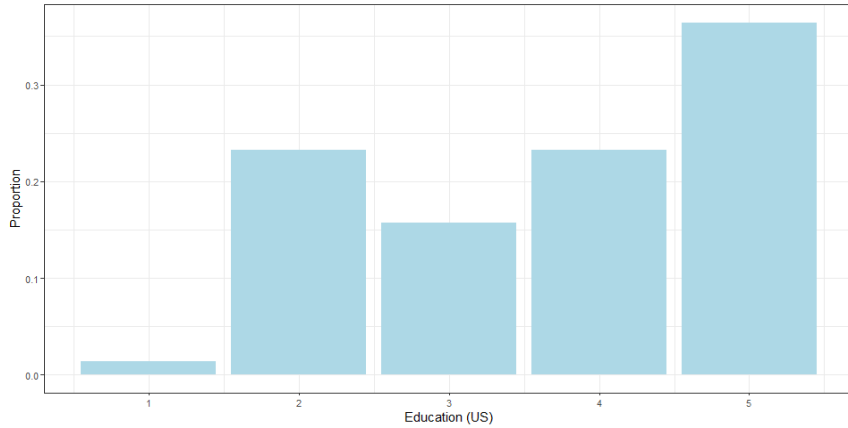
The Erasmus logo, featuring the word "Erasmus" in a stylized, cursive script.

Demographics – Bad data sample



Demographics – Country-specific (bad data)

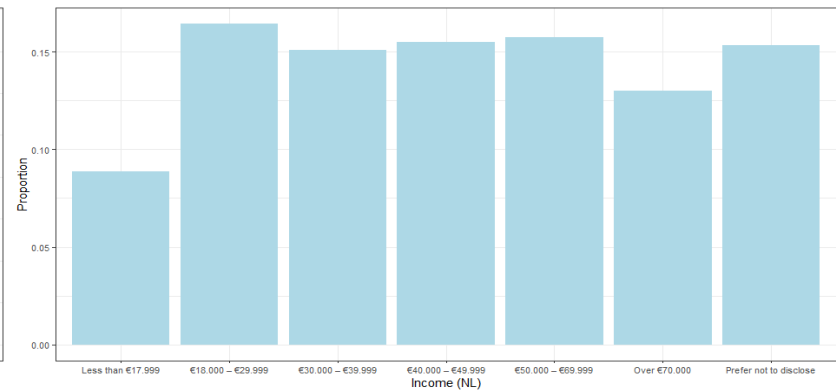
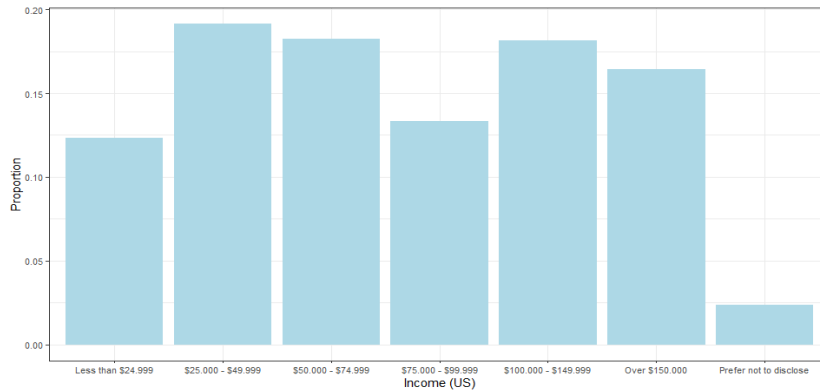
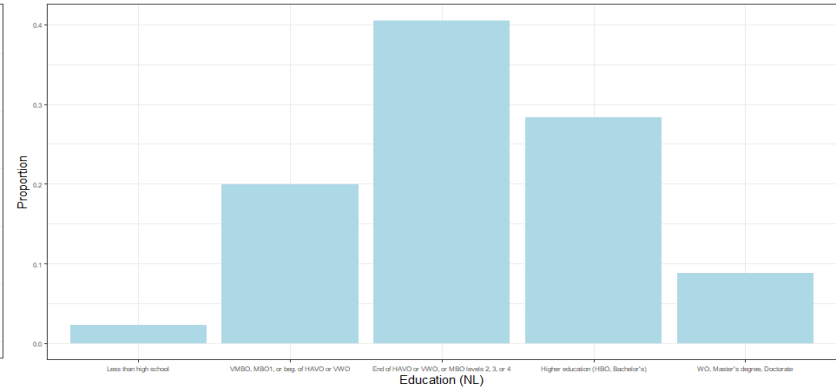
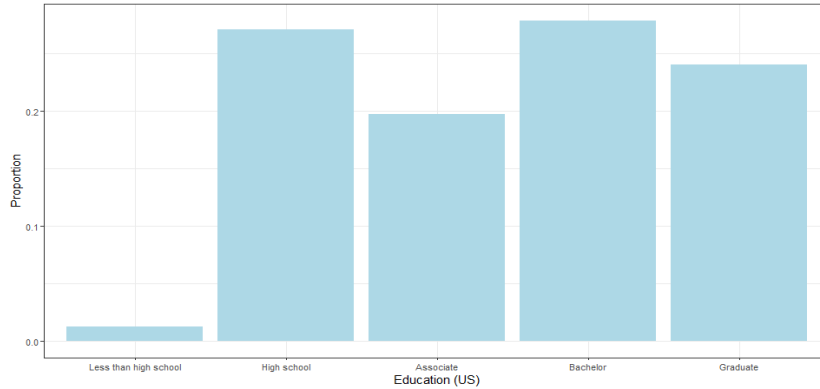
Country-specific demographics from bad data sample



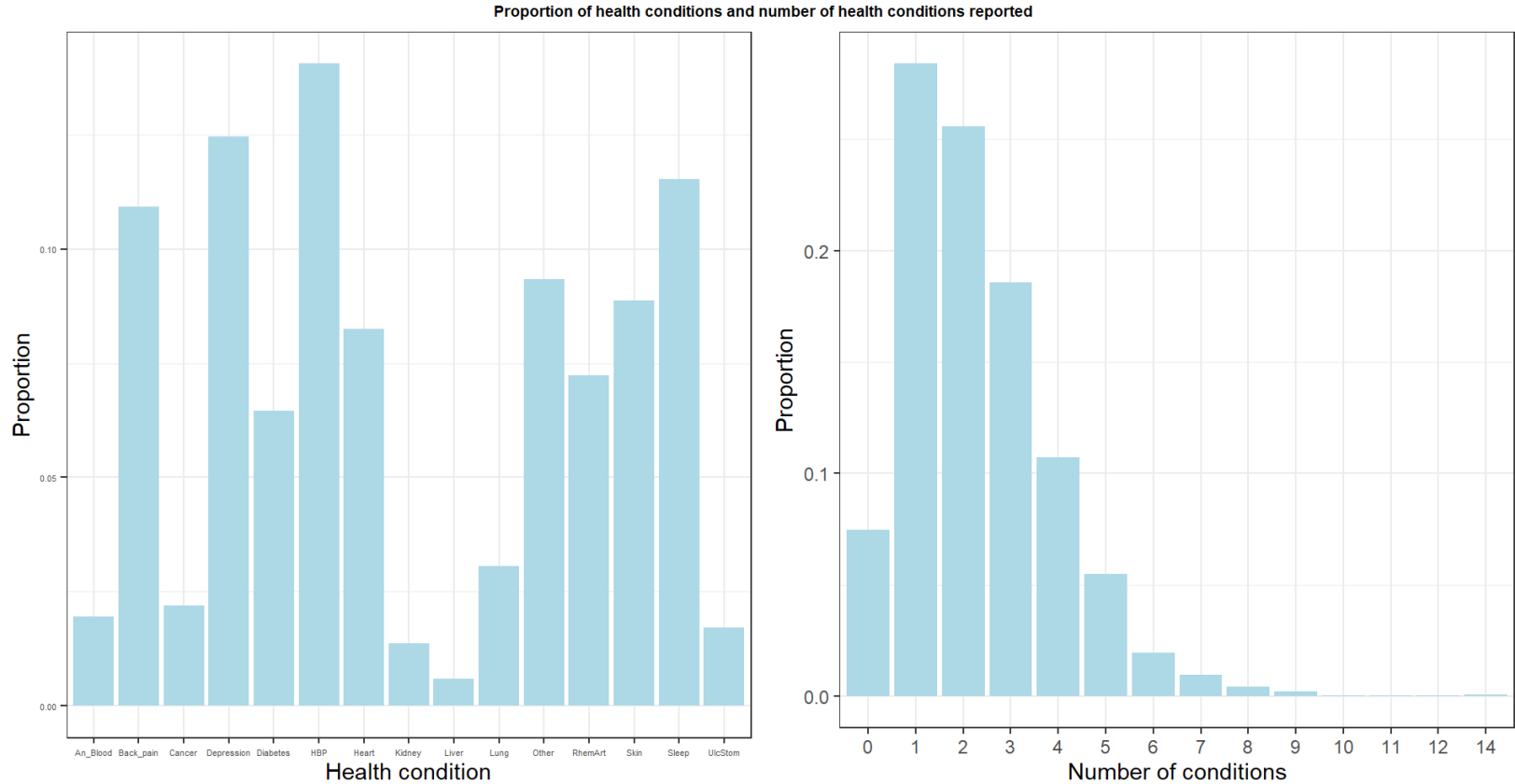
Demographics – Country-specific

- 7155 final sample (US: n= 3606, 50%; NL: n=3549, 50%)

Country-specific demographics from final sample

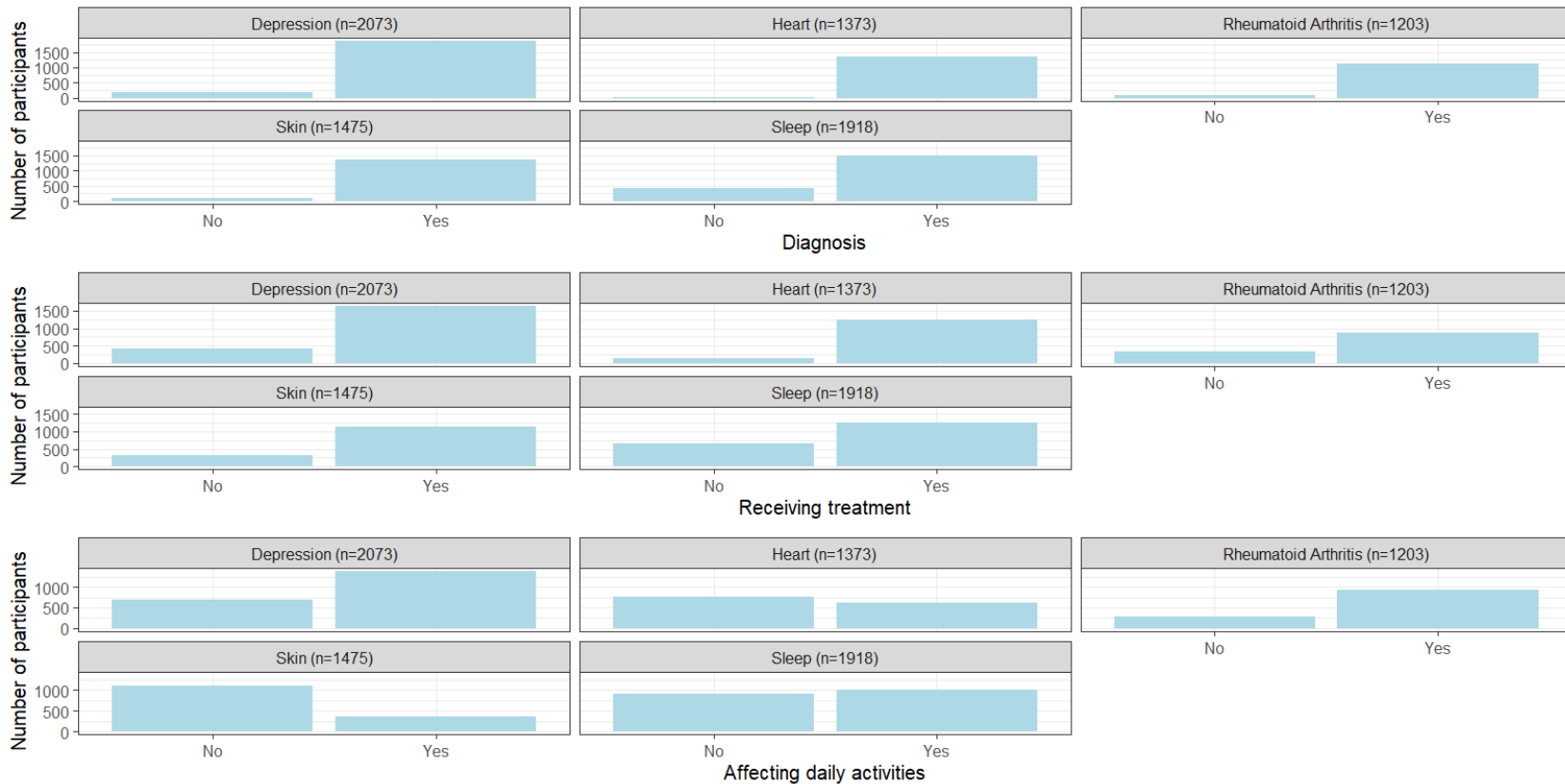


Health conditions



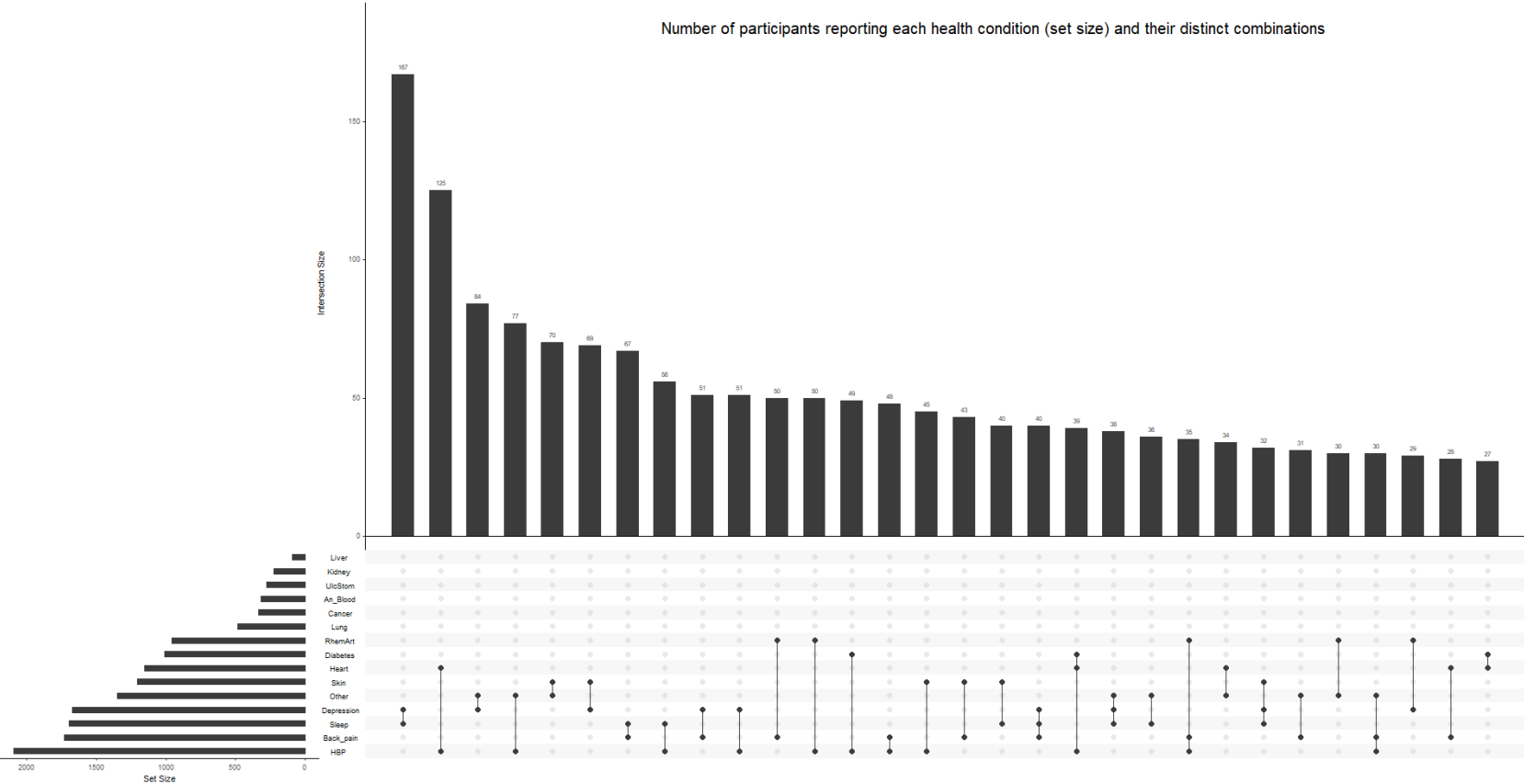
Targeted health conditions

Number of participants with self-reported diagnosis, treatment and being affected in their daily activities



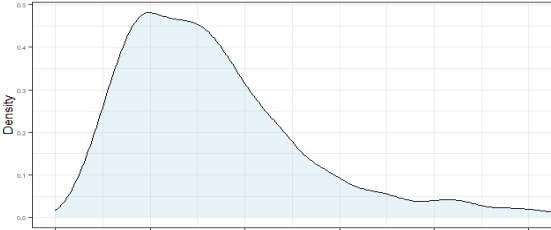
Comorbidity

Number of participants reporting each health condition (set size) and their distinct combinations

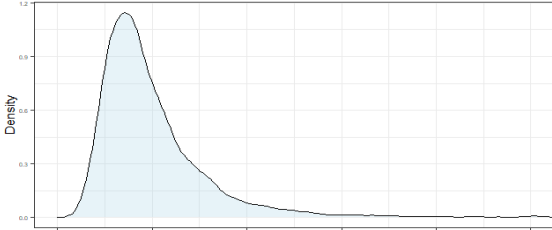


Completion time

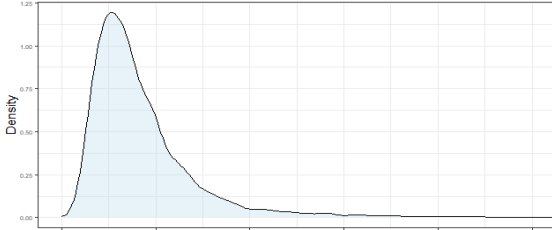
Time completion across instruments



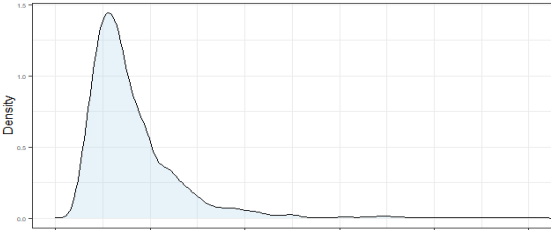
Completion time (ASCOT)



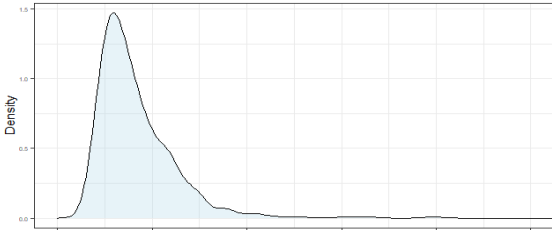
Completion time (EQ-HWB-9)



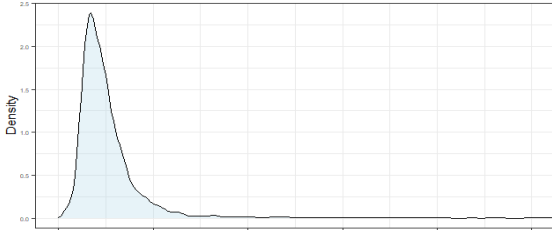
Completion time (ICECAP-A)



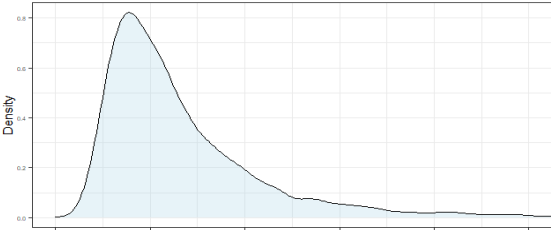
Completion time (ICECAP-O)



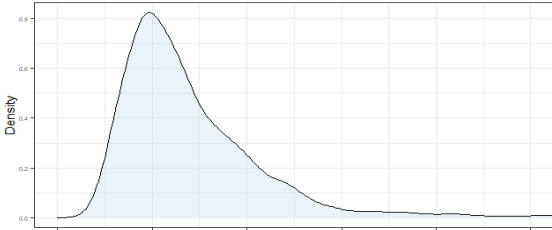
Completion time (QOL-ACC)



Completion time (SWB-5D)



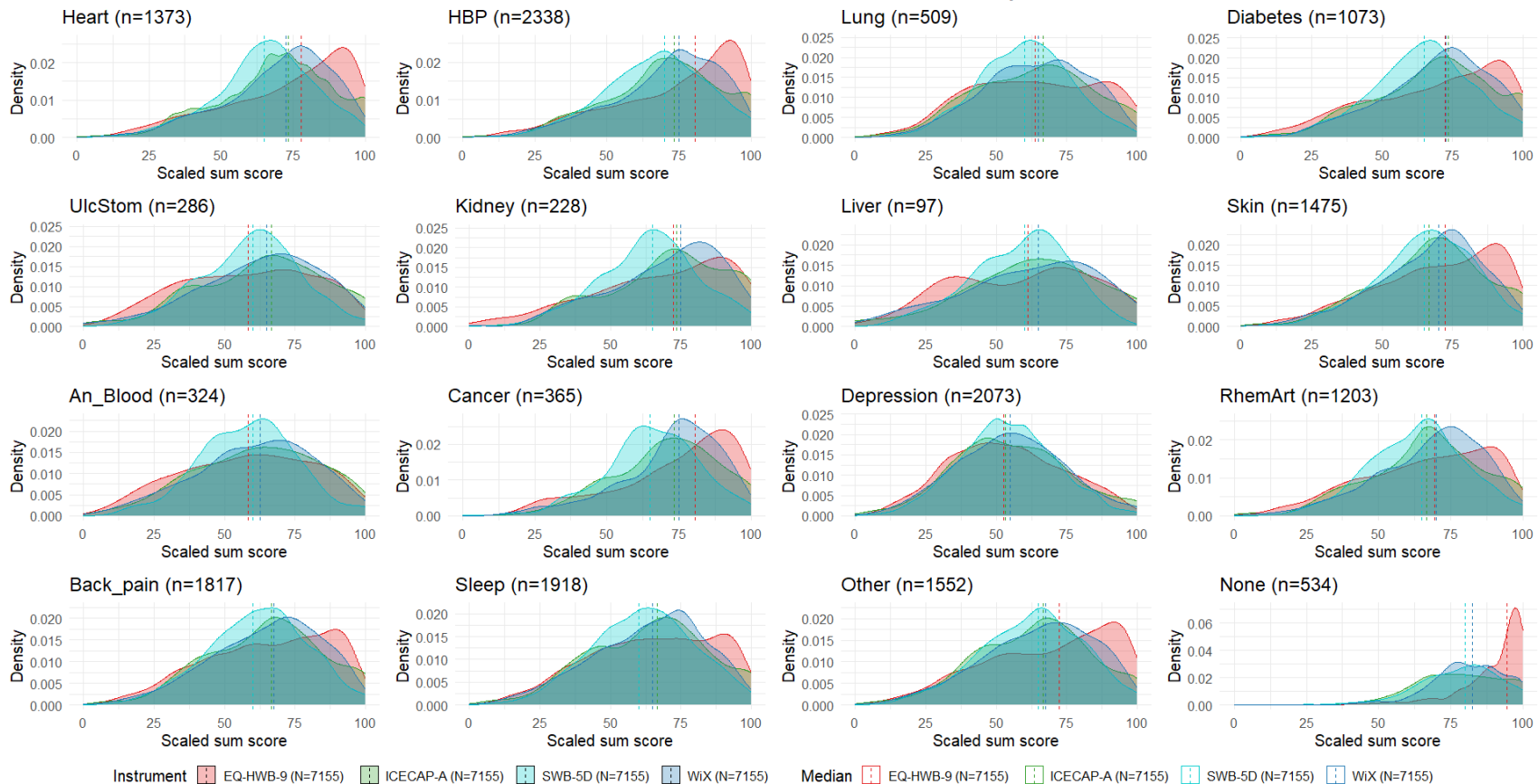
Completion time (WIX)



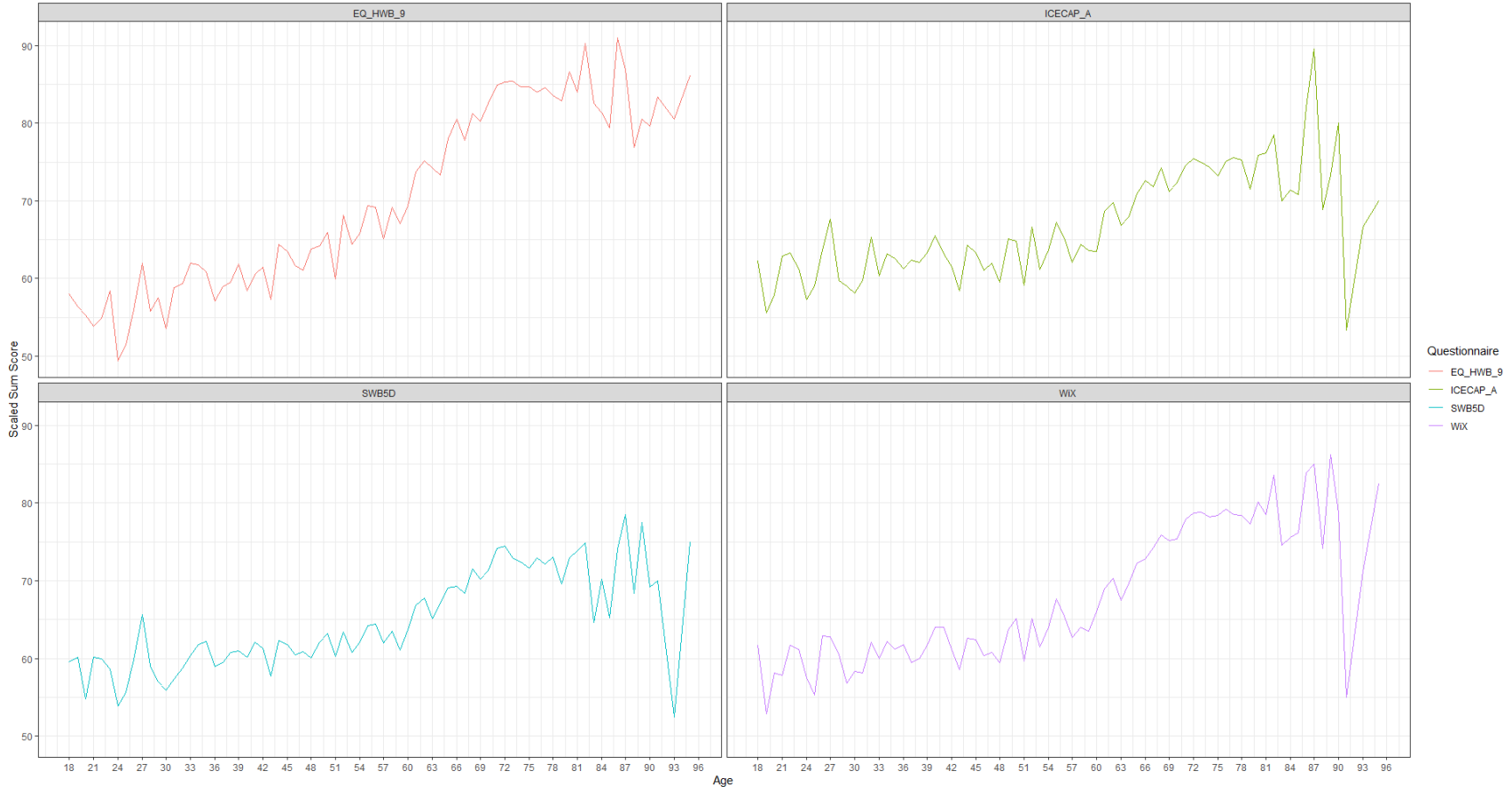
Completion time (WOOP)

Descriptive statistics – Health conditions

Distribution of scaled sum scores of the EQ-HWB-9, ICECAP-A, SWB-5D and WiX by health condition

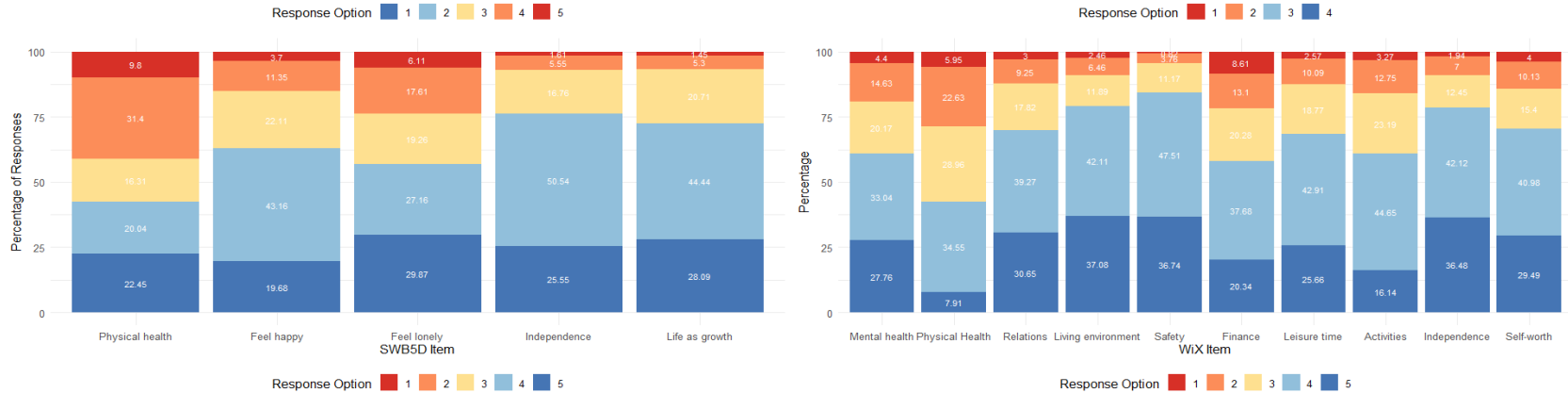
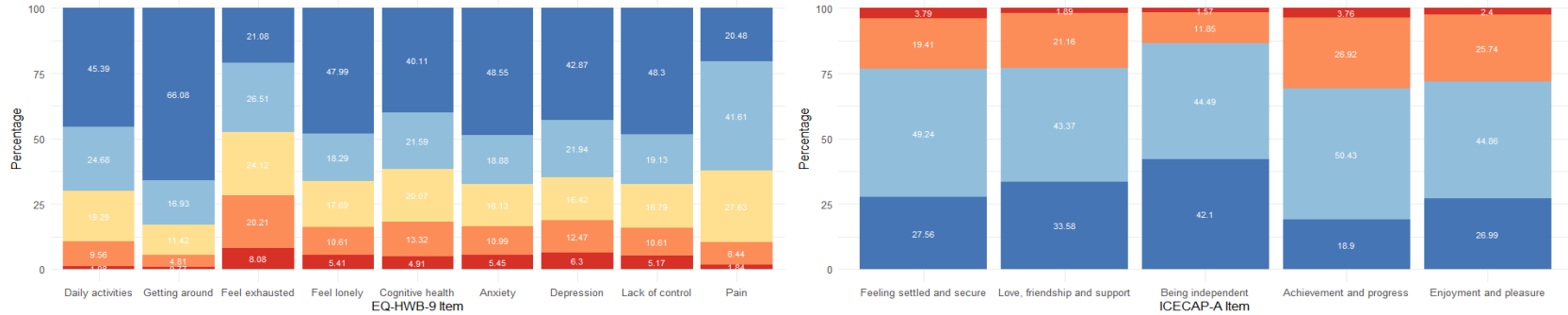


Descriptive statistics – Scores by age



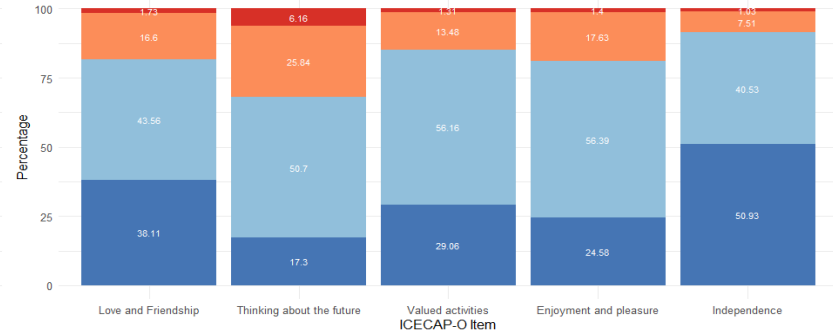
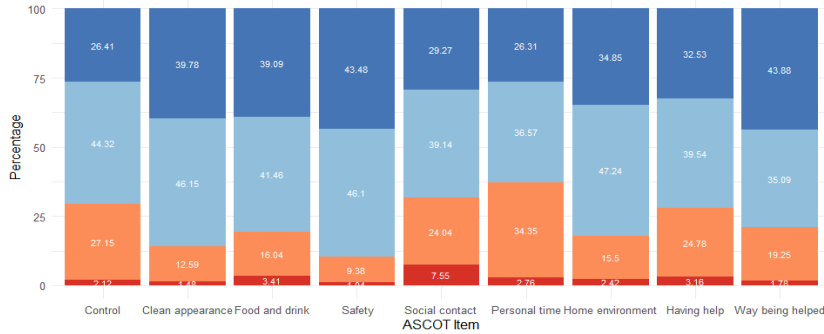
Response options frequencies - Main

Distribution of response option frequencies across well-being instruments



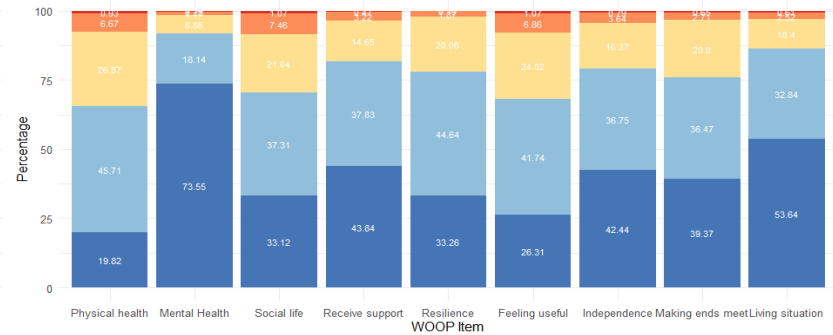
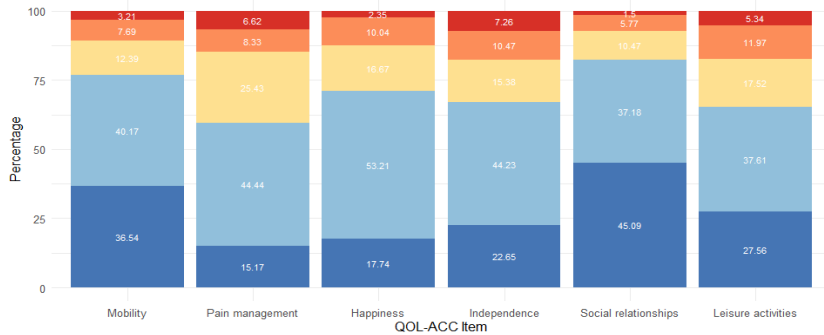
Response options frequencies - Other

Distribution of response option frequencies across well-being instruments



Response Option 1 2 3 4

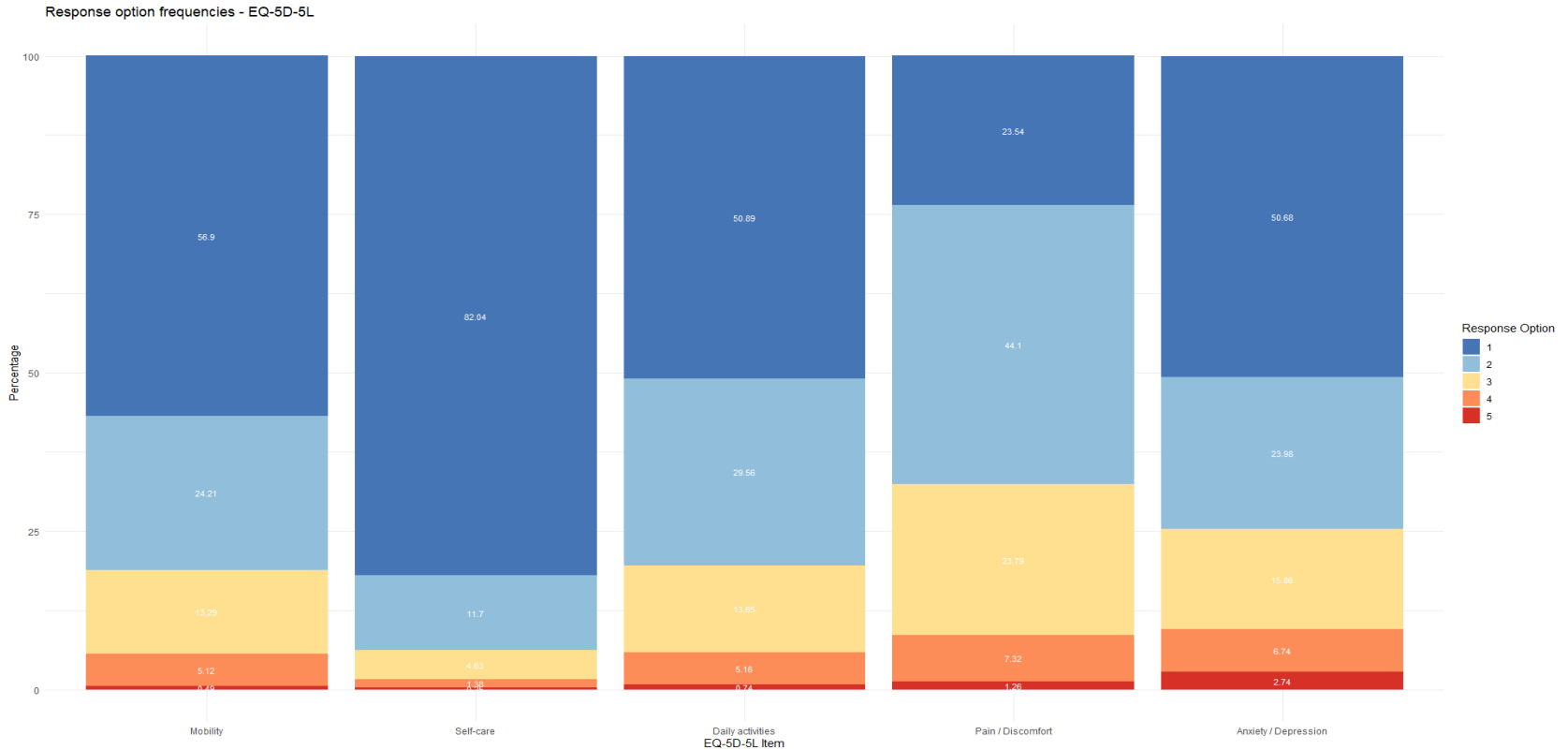
Response Option 1 2 3 4



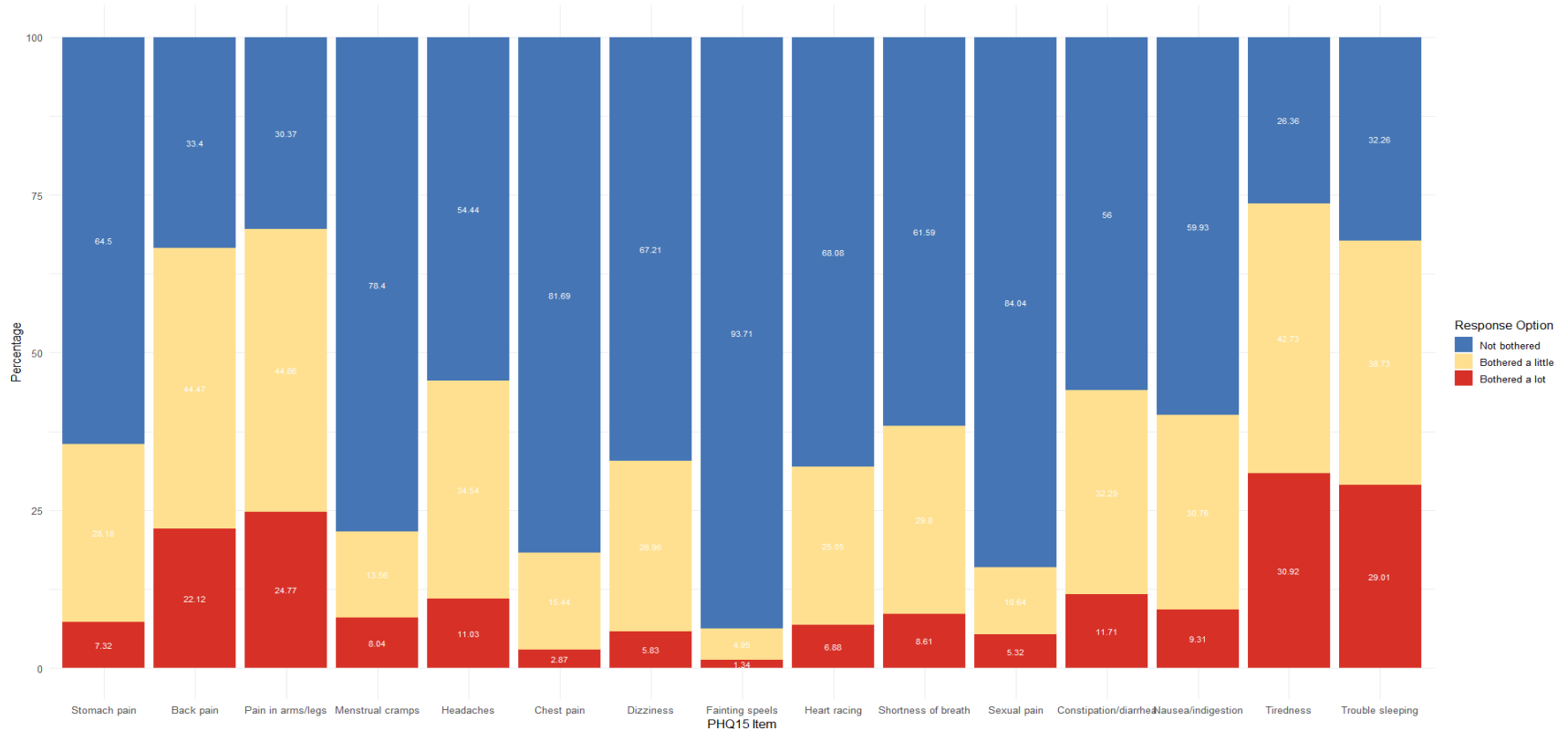
Response Option 0 1 2 3 4

Response Option 0 1 2 3 4 5

Response options frequencies – EQ-5D

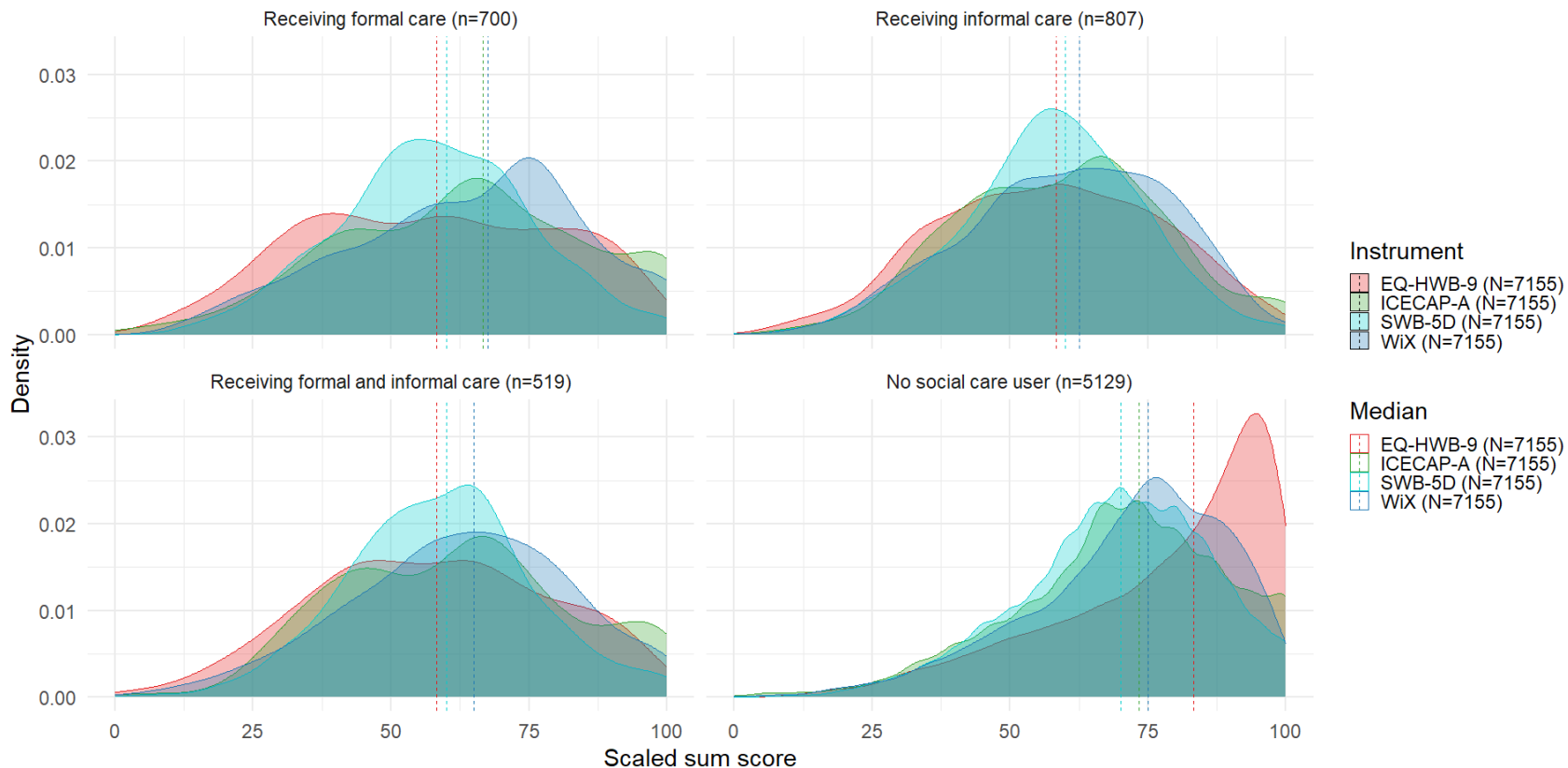


Response options frequencies – PHQ-15



Descriptive statistics – Social care users

Distribution of scaled sum scores of the EQ-HWB-9, ICECAP-A, SWB-5D and WiX by type of social care user



Known-group validity

- All well-being instruments were able to distinguish between demographics groups, to a lower extent

Group	EQ-HWB-9	ICECAP-A	SWB-5D	WiX
Age				
18-24 (n=207) – Mean (SD)	54.82 (20.68)	60.06 (18.83)	58.04 (17.92)	58.71 (17.77)
25-34 (n=575) – Mean (SD)	58.13 (21.67)	61.52 (21.27)	59.12 (18.17)	59.90 (19.84)
35-44 (n=937) – Mean (SD)	60.16 (21.80)	62.61 (21.15)	60.64 (17.22)	61.54 (20.18)
45-54 (n=1057) – Mean (SD)	64.02 (22.32)	62.72 (20.97)	61.57 (17.34)	62.37 (19.24)
55-69 (n=2235) – Mean (SD)	73.99 (20.54)	68.36 (19.14)	66.56 (17.34)	69.50 (17.80)
70+ (n=2144) – Mean (SD)	84.44 (15.34)	74.47 (17.26)	72.57 (15.89)	78.23 (14.28)
Kruskal-Wallis results – H (p) [η^2]	1525.026 (<0.001) [0.21]	450.4962 (<0.001) [0.06]	591.3893 (<0.001) [0.08]	1011.029 (<0.001) [0.14]
Gender				
Female (n=3944)	68.94 (22.01)	65.72 (20.02)	64.60 (17.90)	66.58 (18.98)
Male (n=3206)	75.85 (21.50)	70.45 (19.56)	67.77 (17.30)	71.89 (18.34)
Mann-Whitney U results – U (p) [r]	5060303 (<0.001) [0.17]	5422128 (<0.001) [0.12]	5624552 (<0.001) [0.10]	5213512 (<0.001) [0.15]
Income				
Low (n=2035)	63.32 (23.34)	60.11 (21.40)	60.41 (18.47)	60.68 (20.32)
Medium (n=2223)	73.52 (21.21)	68.38 (19.05)	66.55 (17.34)	69.41 (18.04)
High (n=2267)	77.34 (19.95)	74.17 (17.64)	70.38 (16.36)	75.53 (16.30)
Kruskal-Wallis results – H (p) [η^2]	431.3689 (<0.001) [0.07]	509.4986 (<0.001) [0.08]	331.9955 (<0.001) [0.05]	630.9963 (<0.001) [0.10]
Education				
Lower (n=1814)	67.23 (23.66)	63.00 (21.62)	62.33 (18.09)	64.40 (20.09)
Middle (n=2150)	71.22 (21.48)	65.94 (19.13)	64.71 (17.44)	66.75 (18.22)
Higher (n=3191)	75.25 (20.98)	71.81 (18.75)	68.96 (17.19)	72.98 (17.83)
Kruskal-Wallis results – H (p) [η^2]	147.6623 (<0.001) [0.02]	250.6048 (<0.001) [0.03]	177.7297 (<0.001) [0.02]	299.6006 (<0.001) [0.04]
Relationship				
Not in a relationship (n=2510)	67.31 (22.97)	62.37 (20.84)	62.03 (18.44)	63.35 (19.93)
In a relationship (n=4645)	74.54 (21.15)	70.76 (18.86)	68.15 (16.93)	71.95 (17.62)
Mann-Whitney U results – U (p) [r]	4737502 (<0.001) [0.16]	4468296 (<0.001) [0.19]	4692132 (<0.001) [0.16]	4340843 (<0.001) [0.21]

Effect size interpretation: small ($0.01 \leq \eta^2 < 0.06$, $0.1 \leq r < 0.24$); moderate ($.06 \leq \eta^2 < 0.14$, $0.24 \leq r < 0.37$); strong ($\eta^2 \geq 0.14$, $r \geq 0.37$)

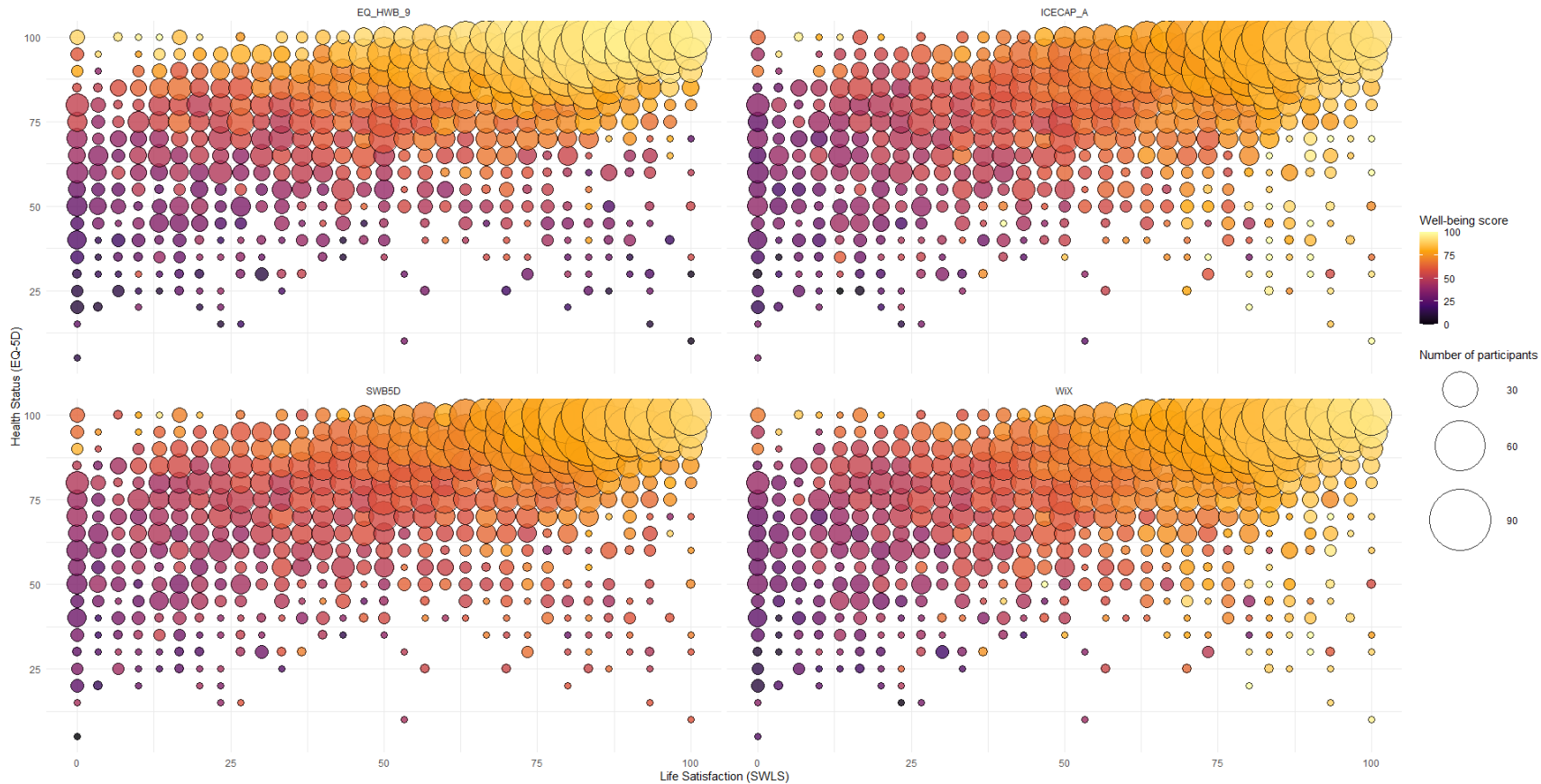
Responsiveness

- None of the instruments was responsive to change

Instrument	Baseline score – Mean (SD)	Follow-up score – Mean (SD)	Score change - Mean	p-value (Effect size)
Change in health				
EQ-HWB-9				
Improved (n=82)	72.66 (21.66)	76.15 (19.51)	3.49	0.33 (N/A)
Worsened (n=96)	57.20 (25.10)	54.46 (23.11)	-2.74	0.43 (N/A)
ICECAP-A				
Improved (n=82)	70.16 (18.66)	72.28 (15.87)	2.12	0.59 (N/A)
Worsened (n=96)	56.18 (21.92)	53.06 (20.43)	-3.12	0.30 (N/A)
SWB-5D				
Improved (n=82)	68.48 (16.15)	70.85 (13.38)	2.37	0.41 (N/A)
Worsened (n=96)	54.32 (18.95)	52.92 (16.59)	-1.40	0.59 (N/A)
WiX				
Improved (n=82)	72.04 (17.24)	74.60 (15.80)	2.56	0.36 (N/A)
Worsened (n=96)	55.57 (21.32)	53.91 (20.67)	-1.66	0.48 (N/A)
Change in well-being				
EQ-HWB-9				
Improved (n=113)	77.31 (21.57)	80.06 (19.47)	2.75	0.51 (N/A)
Worsened (n=103)	53.34 (24.48)	51.78 (21.65)	-1.56	0.66 (N/A)
ICECAP-A				
Improved (n=113)	74.93 (19.85)	75.87 (16.31)	0.94	0.94 (N/A)
Worsened (n=103)	52.36 (21.13)	49.51 (20.66)	-2.85	0.33 (N/A)
SWB-5D				
Improved (n=113)	71.99 (16.39)	75.04 (12.72)	3.05	0.25 (N/A)
Worsened (n=103)	52.72 (20.84)	51.75 (17.56)	-0.97	0.74 (N/A)
WiX				
Improved (n=113)	75.20 (19.05)	78.38 (14.84)	3.18	0.35 (N/A)
Worsened (n=103)	53.06 (19.98)	51.75 (19.26)	-1.31	0.61 (N/A)

Instrument mapping

Mapping of instrument scores across health status (EQ-5D) and life satisfaction (SWLS)



Previous literature

- EQ-HWB-9 did not detect significant differences between caregivers and the general population (1)
 - Perhaps the EQ-HWB-9 may have lost some discriminatory power in broader well-being compared to the full EQ-HWB
- EQ-HWB and ICECAP-A had to distinct factor structures (2)
 - EQ-HWB items loading on physical functioning and psychosocial health factors
 - ICECAP-A loading on a separate well-being factor
- Recent reviews concluded the ICECAP-A could offer a complementary (3) or even enhanced assessment of well-being (4) if targeting broader HRQOL

1) Kuharić M, Pickard AS, Mukuria C, Finch AP. The Measurement Properties of the EQ-HWB and the EQ-HWB-S in Italian Population: A Comparative Study With EQ-5D-5L. *Value Health*. 2024 Jul;27(7):955–66.

2) Kinchin I, Engel L, Rencz F. A Comparative Study of Health and Well-Being Measures in Ireland Using EQ Health and Well-Being (EQ-HWB) and its Short Version, EQ-5D-5L, and ICEpop Capability Measure for Adults (ICECAP-A). *Value Health*. 2025 May;S1098301525023241.

3) Afentou N, Kinghorn P. A Systematic Review of the Feasibility and Psychometric Properties of the ICEpop CAPability Measure for Adults and Its Use So Far in Economic Evaluation. *Value Health*. 2020 Apr;23(4):515–26.

4) Heltter, T. M. (2020). *Capability instruments in economic evaluations of health-related interventions: a comparative review of the literature*. *Quality of Life Research*.

The Erasmus logo is a stylized, handwritten-style signature of the name 'Erasmus' in black ink, located in the bottom right corner of the slide.