HARMONY

Transforming data management with the Harmony Data Tool: A hands-on introduction

UK Data Service Workshop
29th Nov 2024

Dr Eoin McElroy
School of Psychology
Ulster University

Thomas Wood
Fast Data Science

wellcome

UKRI Economic and Social Research Council

MEET THE HARMONY TEAM

Rachel Gomes
University College London

Richard Thomas
UK-LLC

Eoin McElroy
Ulster University

Bettina Moltrecht
University College London

Louise Arseneault
Kings College London

Thomas Wood
Fast Data Science

Mauricio Hoffmann
Universidade Federal de Santa Maria

CENTRE FOR LONGITUDINAL STUDIES

Ulster University

UK Longitudinal Linkage Collaboration

Fast Data Science

Outline

- Overview of data harmonisation (with a focus on longitudinal data)
- Introduction to Harmony
- **Demo 1:** Hands-on demonstration of the web-based version
- **Demo 2:** Hands-on demonstration of the R version
- **Demo 3:** Hands-on demonstration of the Python version
- **Demo 4:** Hands-on demonstration of the API
- Showcasing different use cases and integrations

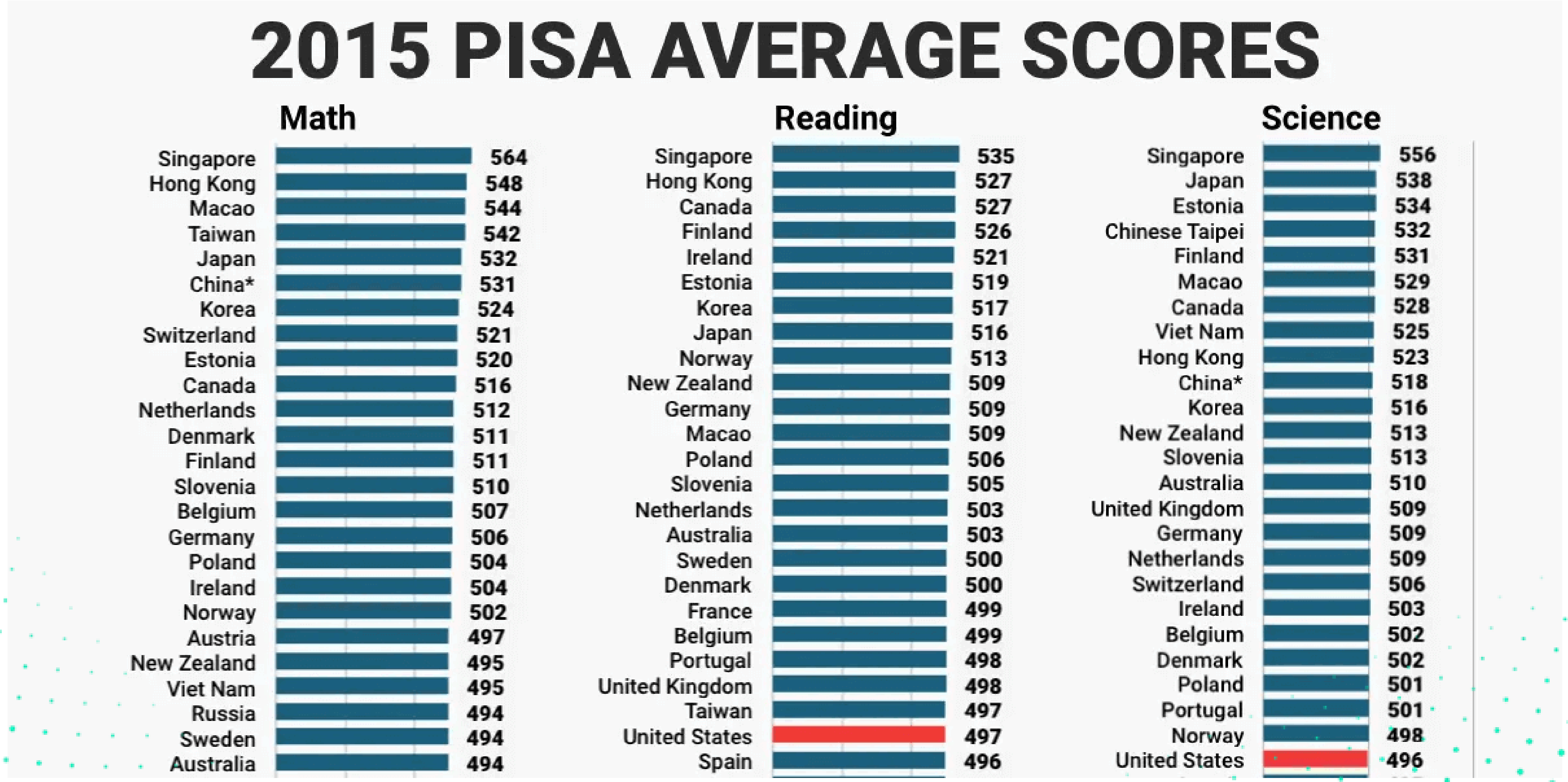
- Interactive Q&A

Data Harmonisation

“A process that aims to produce equivalent or comparable measures of a given characteristic across datasets coming from different populations or from the same population but at different time points”

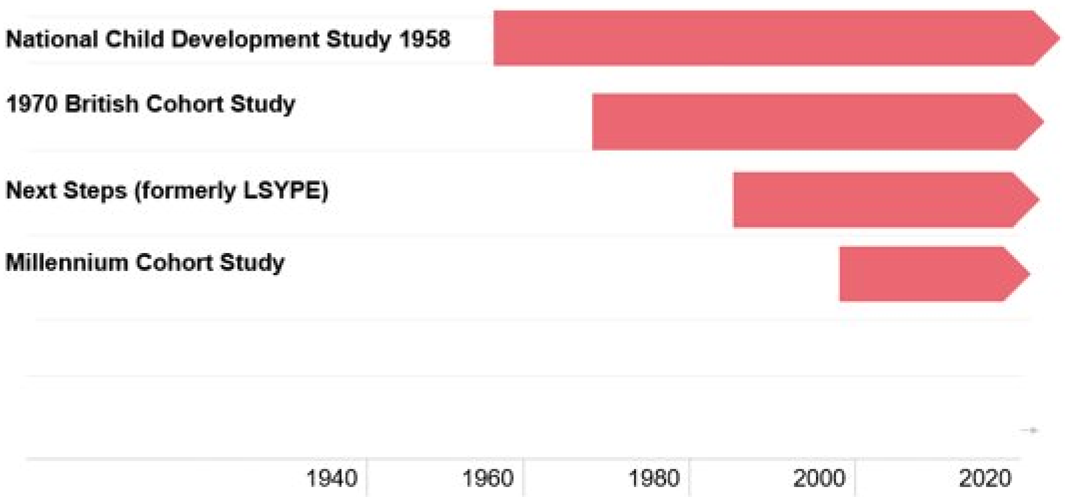
(Tomescu-Dubrow et al., 2024)

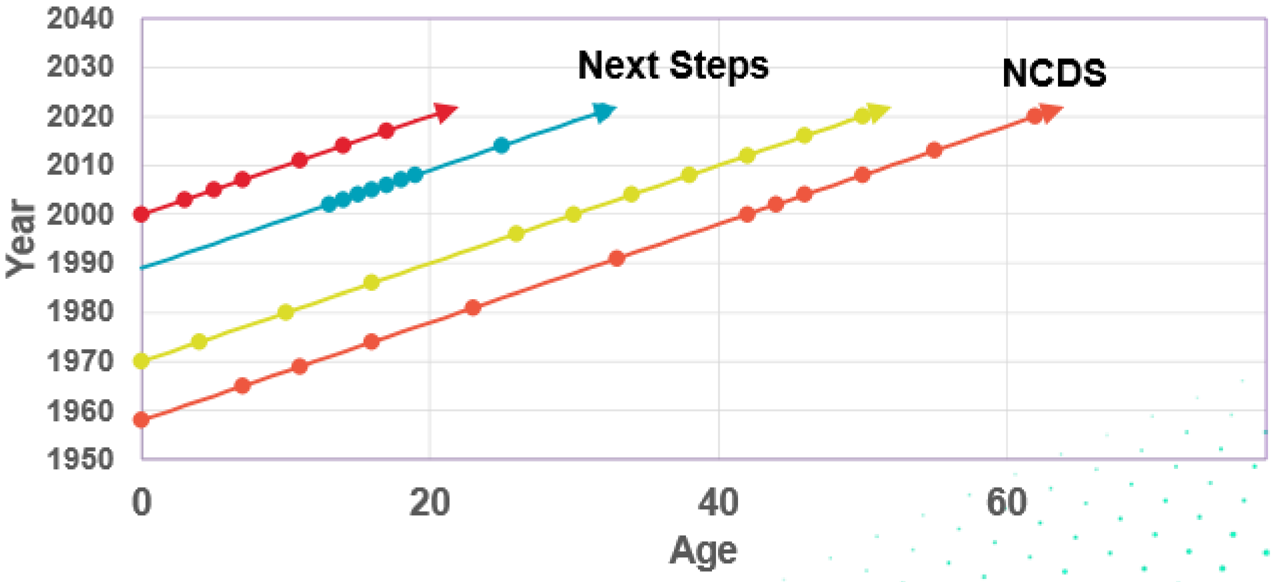
Data Harmonisation
- prospective vs retrospective



Data Harmonisation

- prospective vs retrospective





CENTRE FOR LONGITUDINAL STUDIES

UK Data Service

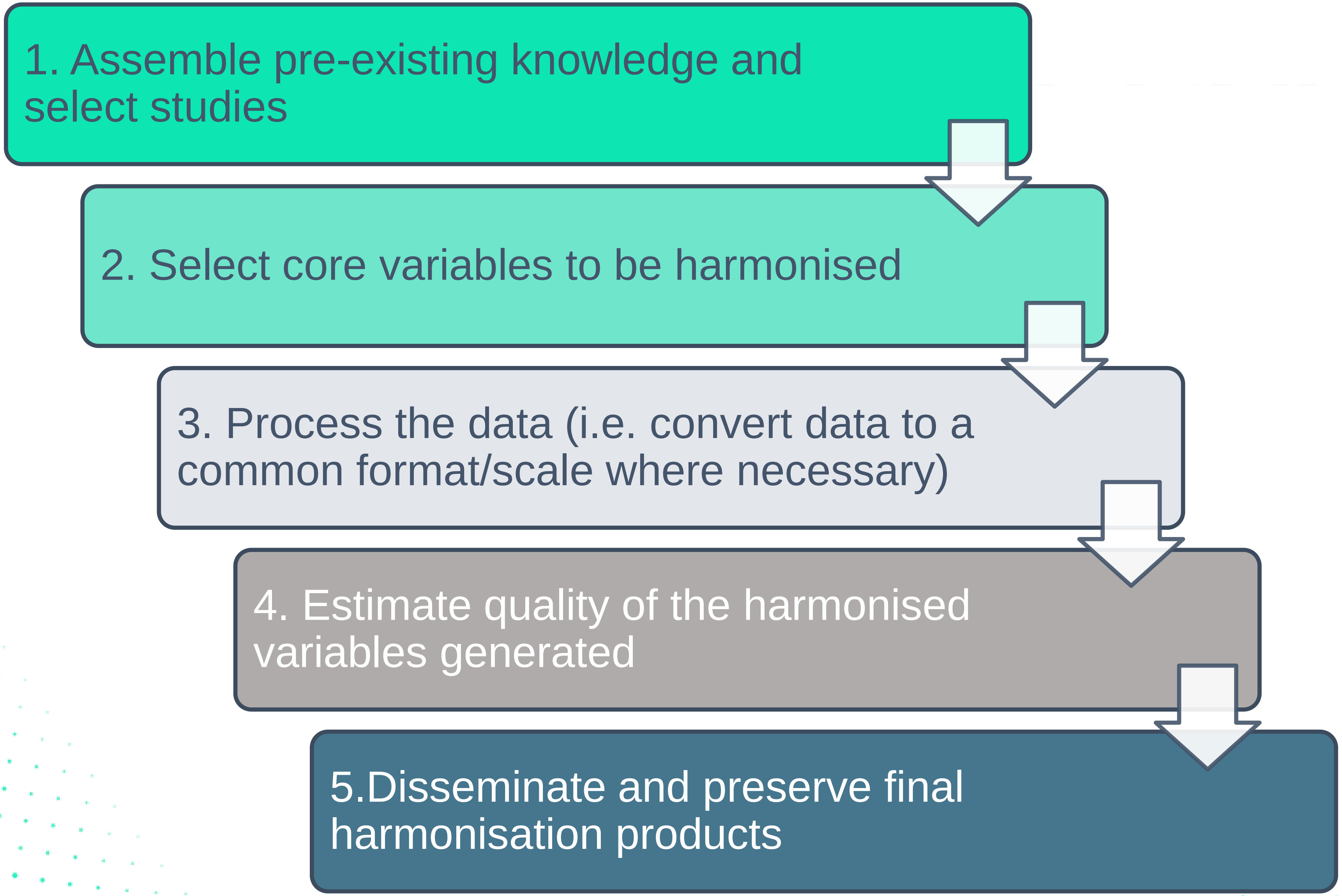
Harmonisation – benefits

Harmonisation – benefits

- Findable
- Accessible
- Interoperable
- Reusable

Harmonisation – challenges

- Availability
- Comparability vs equivalence (Tomescu-Dubrow et al., 2024)
- Loss of information
- Processor degrees of freedom



1. Assemble pre-existing knowledge and select studies

2. Select core variables to be harmonised

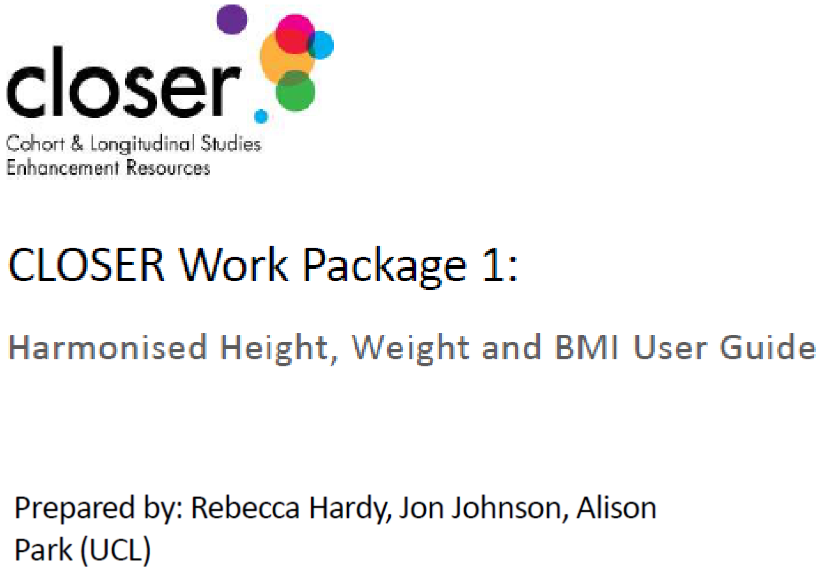
3. Process the data (i.e. convert data to a common format/scale where necessary)

4. Estimate quality of the harmonised variables generated

5.Disseminate and preserve final harmonisation products

(Fortier et al., 2017)

Harmonisation – types of data



closer

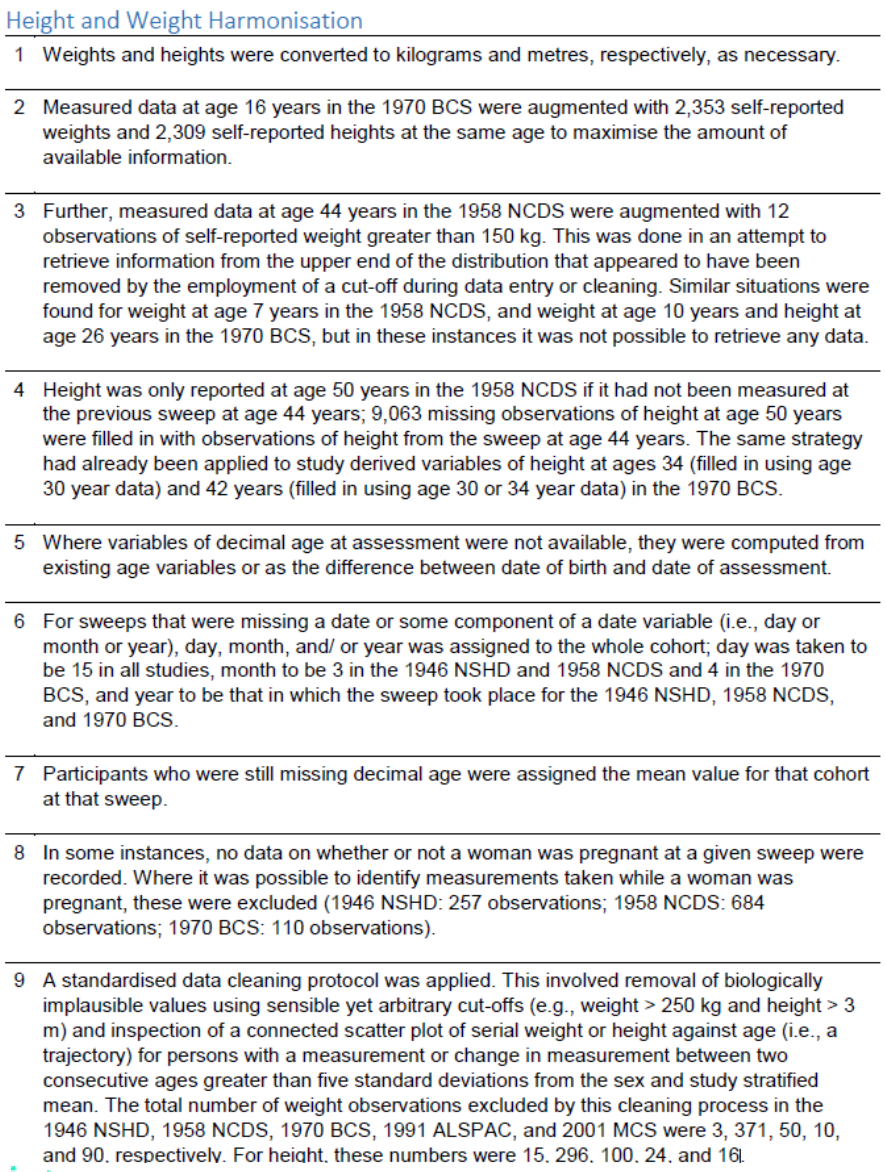
Cohort & Longitudinal Studies
Enhancement Resources

CLOSER Work Package 1:

Harmonised Height, Weight and BMI User Guide

Prepared by: Rebecca Hardy, Jon Johnson, Alison
Park (UCL)

Cohort and Longitudinal Studies Enhancement Resources. (2017). *Harmonised Height, Weight and BMI in Five Longitudinal Cohort Studies: National Child Development Study, 1970 British Cohort Study and Millennium Cohort Study*. [data collection]. UK Data Service. SN: 8207, DOI: http://doi.org/10.5255/UKDA-SN-8207-1



Height and Weight Harmonisation

1 Weights and heights were converted to kilograms and metres, respectively, as necessary.

2 Measured data at age 16 years in the 1970 BCS were augmented with 2,353 self-reported weights and 2,309 self-reported heights at the same age to maximise the amount of available information.

3 Further, measured data at age 44 years in the 1958 NCDS were augmented with 12 observations of self-reported weight greater than 150 kg. This was done in an attempt to retrieve information from the upper end of the distribution that appeared to have been removed by the employment of a cut-off during data entry or cleaning. Similar situations were found for weight at age 7 years in the 1958 NCDS, and weight at age 10 years and height at age 26 years in the 1970 BCS, but in these instances it was not possible to retrieve any data.

4 Height was only reported at age 50 years in the 1958 NCDS if it had not been measured at the previous sweep at age 44 years; 9,063 missing observations of height at age 50 years were filled in with observations of height from the sweep at age 44 years. The same strategy had already been applied to study derived variables of height at ages 34 (filled in using age 30 year data) and 42 years (filled in using age 30 or 34 year data) in the 1970 BCS.

5 Where variables of decimal age at assessment were not available, they were computed from existing age variables or as the difference between date of birth and date of assessment.

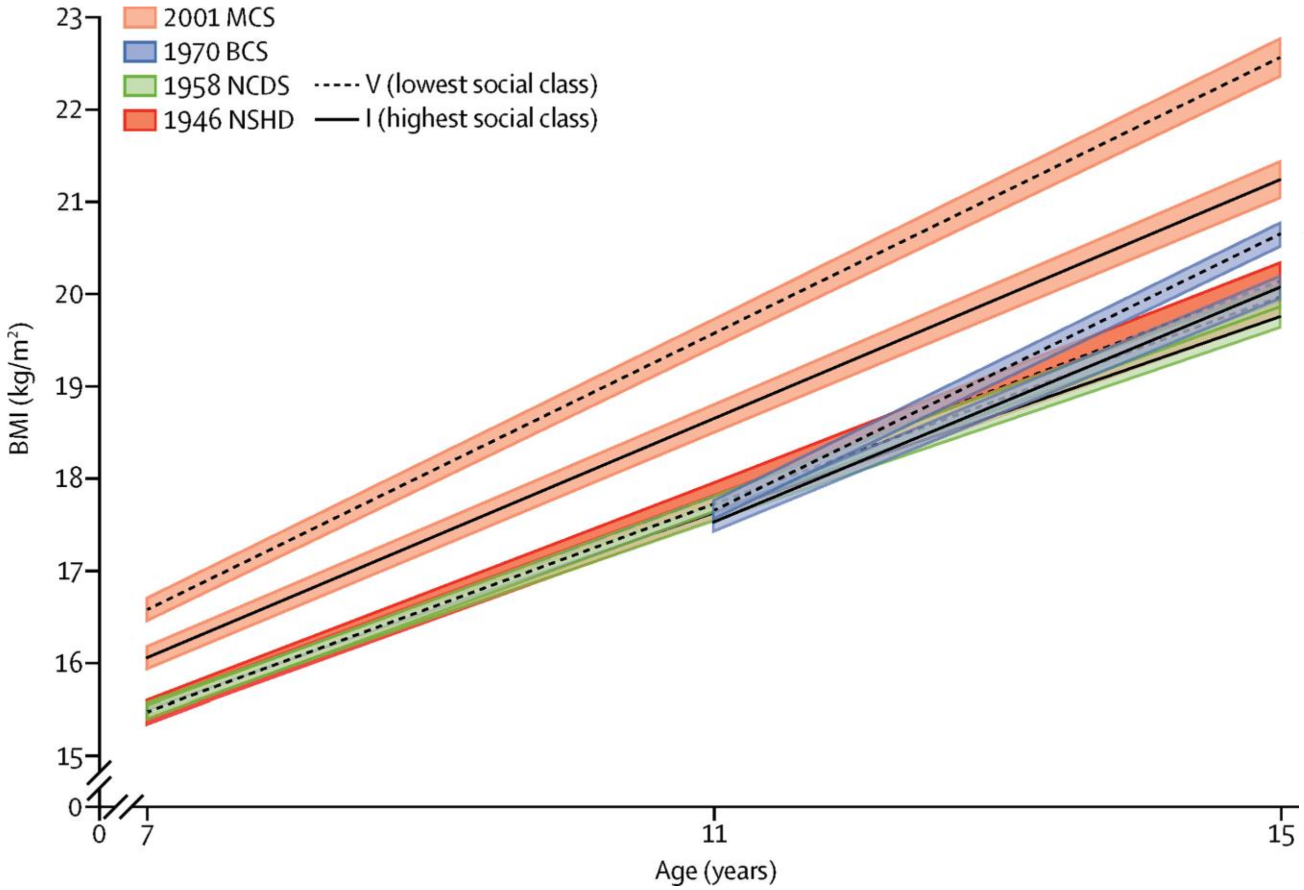
6 For sweeps that were missing a date or some component of a date variable (i.e., day or month or year), day, month, and/ or year was assigned to the whole cohort; day was taken to be 15 in all studies, month to be 3 in the 1946 NSHD and 1958 NCDS and 4 in the 1970 BCS, and year to be that in which the sweep took place for the 1946 NSHD, 1958 NCDS, and 1970 BCS.

7 Participants who were still missing decimal age were assigned the mean value for that cohort at that sweep.

8 In some instances, no data on whether or not a woman was pregnant at a given sweep were recorded. Where it was possible to identify measurements taken while a woman was pregnant, these were excluded (1946 NSHD: 257 observations; 1958 NCDS: 684 observations; 1970 BCS: 110 observations).

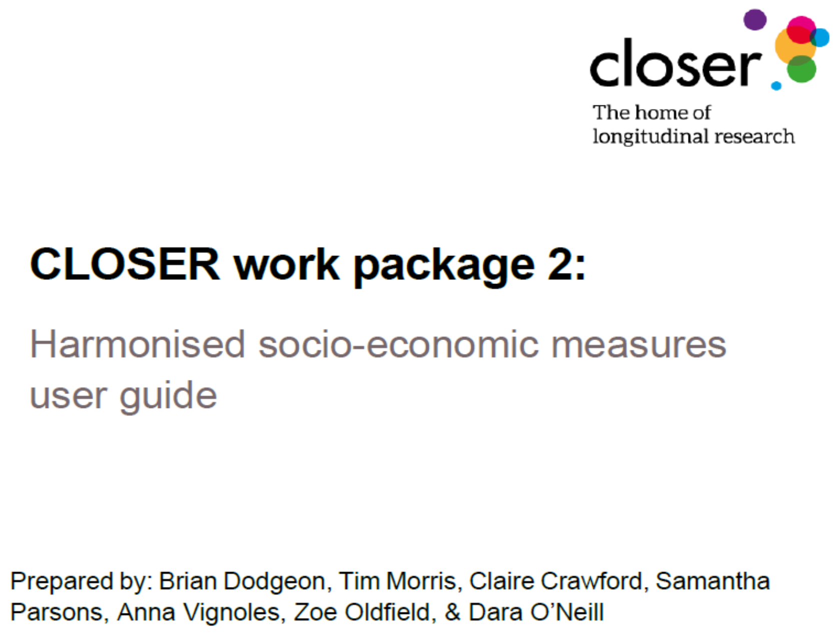
9 A standardised data cleaning protocol was applied. This involved removal of biologically implausible values using sensible yet arbitrary cut-offs (e.g., weight > 250 kg and height > 3 m) and inspection of a connected scatter plot of serial weight or height against age (i.e., a trajectory) for persons with a measurement or change in measurement between two consecutive ages greater than five standard deviations from the sex and study stratified mean. The total number of weight observations excluded by this cleaning process in the 1946 NSHD, 1958 NCDS, 1970 BCS, 1991 ALSPAC, and 2001 MCS were 3, 371, 50, 10, and 90, respectively. For height, these numbers were 15, 296, 100, 24, and 16

http://doi.org/10.5255/UKDA-SN-8207-1



Bann, D., Johnson, W., Li, L., Kuh, D., & Hardy, R. (2018). Socioeconomic inequalities in childhood and adolescent body-mass index, weight, and height from 1953 to 2015: an analysis of four longitudinal, observational, British birth cohort studies. *The Lancet Public Health*, *3*(4), e194-e203.

Harmonisation – types of data



closer
The home of
longitudinal research

CLOSER work package 2:

Harmonised socio-economic measures
user guide

Prepared by: Brian Dodgeon, Tim Morris, Claire Crawford, Samantha
Parsons, Anna Vignoles, Zoe Oldfield, & Dara O'Neill

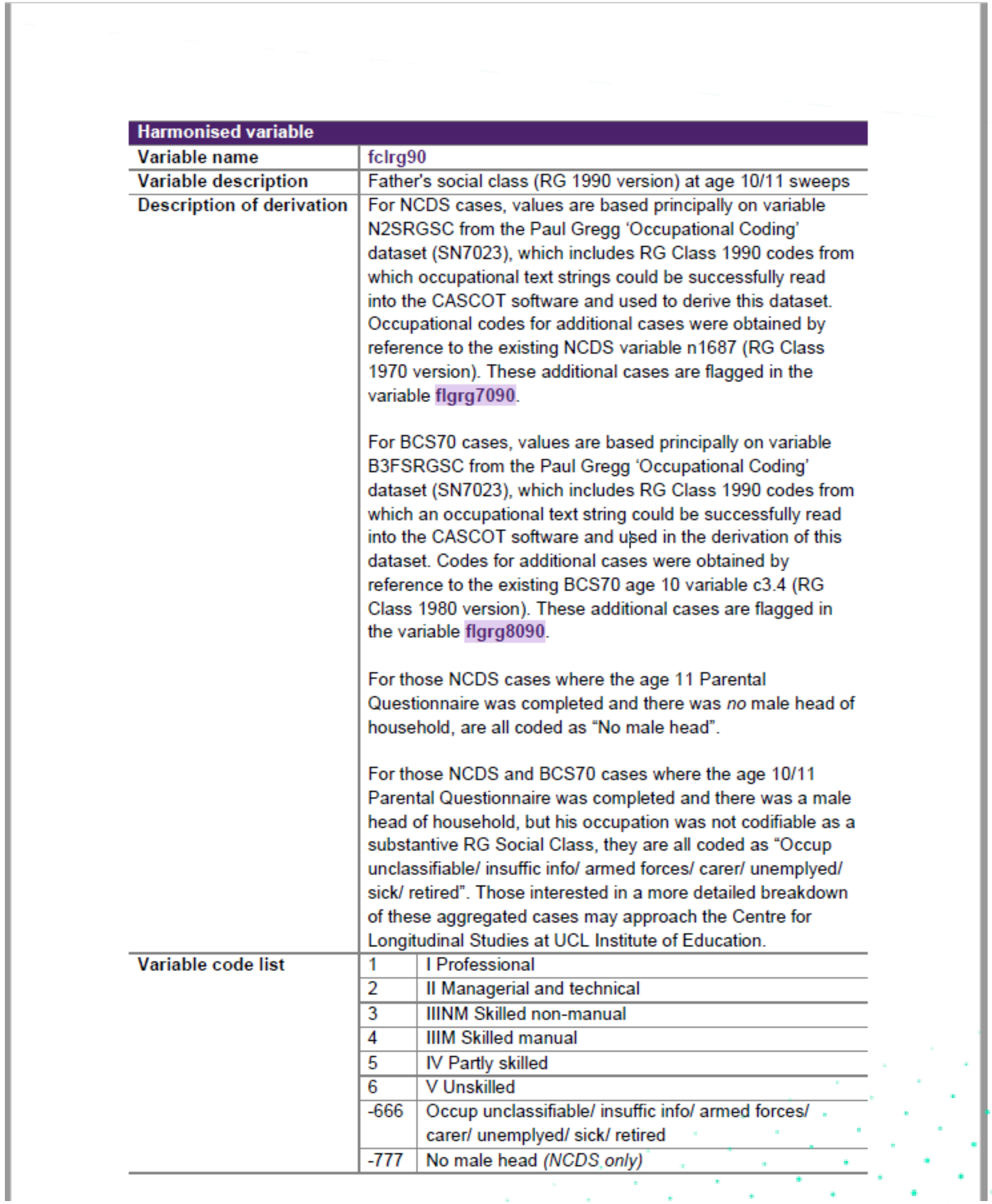
Cohort and Longitudinal Studies Enhancement Resources. (2022). *CLOSER*. [data series]. *8th Release.* UK Data Service. SN: 2000111, DOI: http://doi.org/10.5255/UKDA-Series-2000111

Harmonisation – types of data

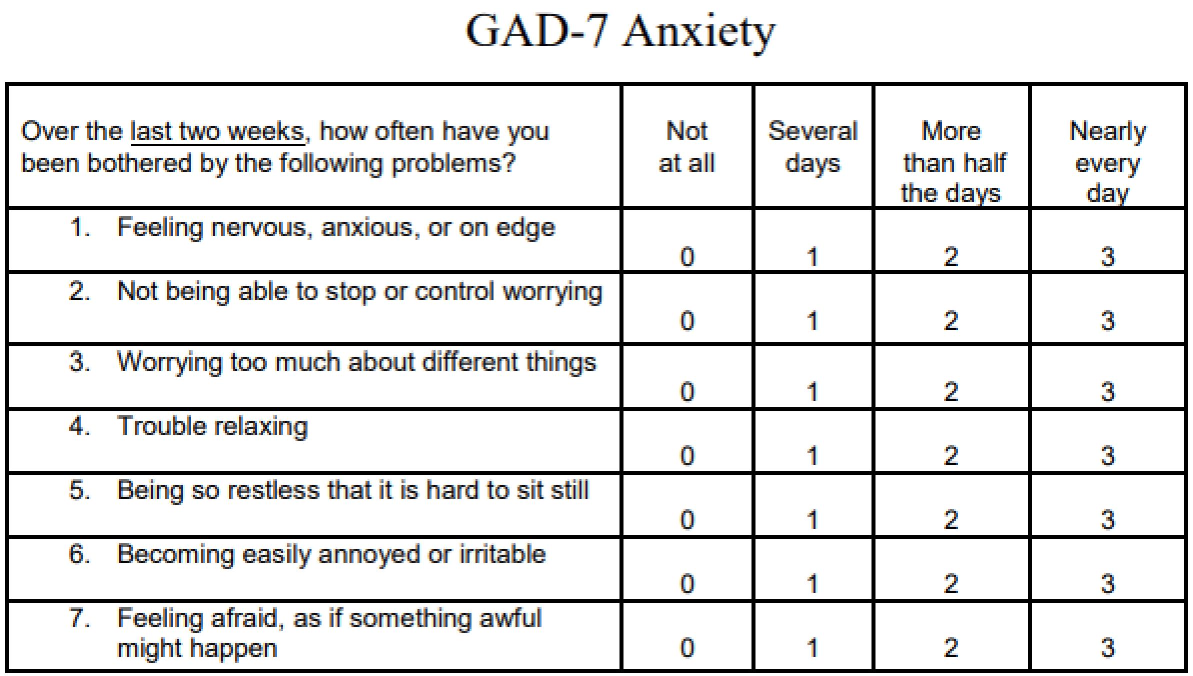
- CO70 (OPCS 1970 Classification Of Occupations)
  - 223 categories, subdivided into 26 Occupational Orders
- CO80 (OPCS 1980 Operational Coding Groups)
  - 350 categories (or 547 ‘Occupational Groups’ when split by supervisory status)
- SOC90
  - 371 ‘unit’ groups, with 77 Minor Groups, 22 Sub-Major Groups, 9 Major Groups
- SOC2000
  - 353 unit groups, 81 Minor Groups, 25 Sub-Major groups, 9 Major Groups
- SOC2010
  - Relatively minor revisions tSOC2000

DOI: http://doi.org/10.5255/UKDA-Series-2000111

DOI: http://doi.org/10.5255/UKDA-Series-2000111

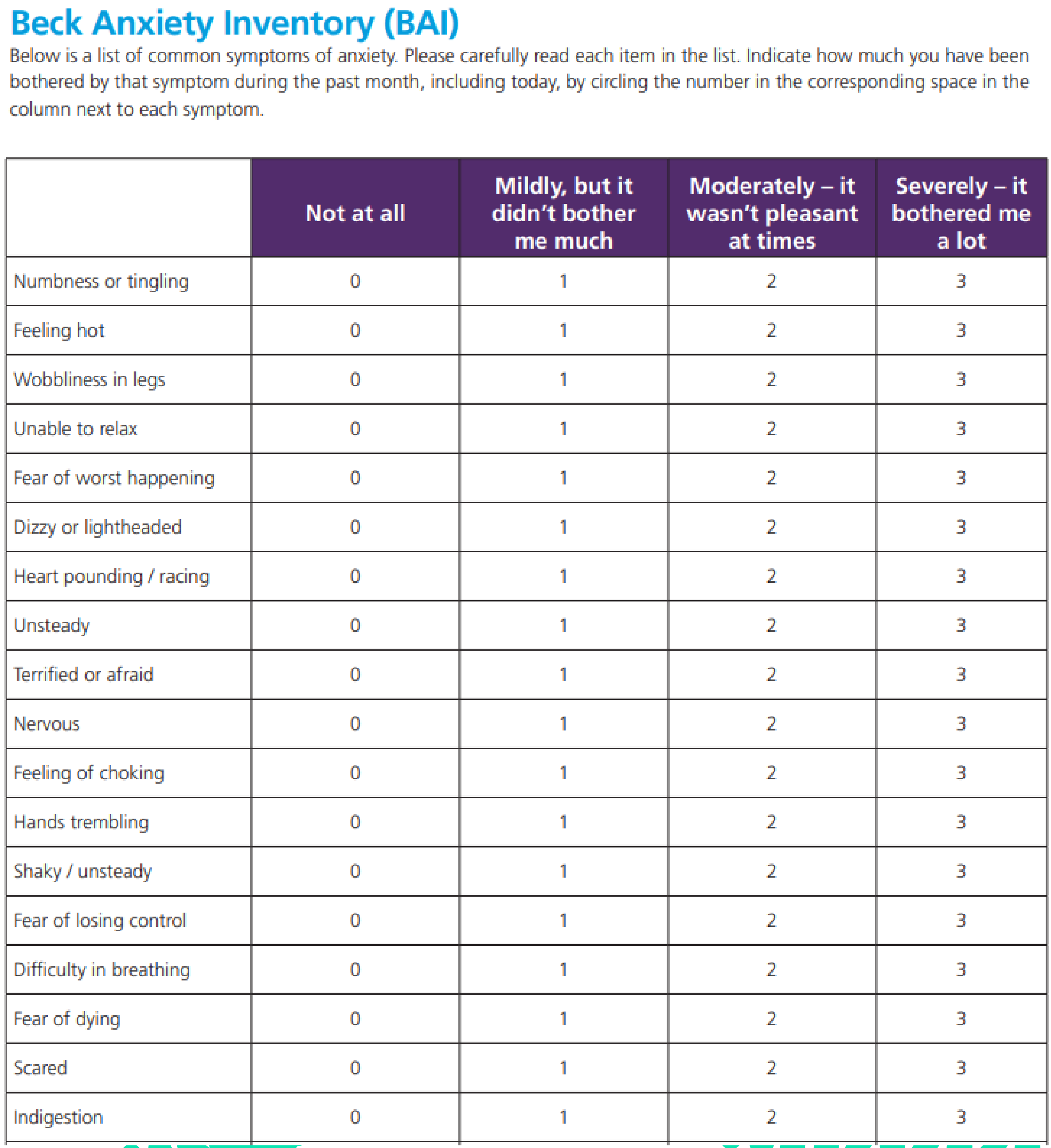


Harmonised variable		
Variable name	fclrg90	
Variable description	Father's social class (RG 1990 version) at age 10/11 sweeps	
Description of derivation	For NCDS cases, values are based principally on variable N2SRGSC from the Paul Gregg 'Occupational Coding' dataset (SN7023), which includes RG Class 1990 codes from which occupational text strings could be successfully read into the CASCOT software and used to derive this dataset. Occupational codes for additional cases were obtained by reference to the existing NCDS variable n1687 (RG Class 1970 version). These additional cases are flagged in the variable flgrg7090. For BCS70 cases, values are based principally on variable B3FSRGSC from the Paul Gregg 'Occupational Coding' dataset (SN7023), which includes RG Class 1990 codes from which an occupational text string could be successfully read into the CASCOT software and used in the derivation of this dataset. Codes for additional cases were obtained by reference to the existing BCS70 age 10 variable c3.4 (RG Class 1980 version). These additional cases are flagged in the variable flgrg8090. For those NCDS cases where the age 11 Parental Questionnaire was completed and there was *no* male head of household, are all coded as "No male head". For those NCDS and BCS70 cases where the age 10/11 Parental Questionnaire was completed and there was a male head of household, but his occupation was not codifiable as a substantive RG Social Class, they are all coded as "Occup unclassifiable/ insuffic info/ armed forces/ carer/ unemplyed/ sick/ retired". Those interested in a more detailed breakdown of these aggregated cases may approach the Centre for Longitudinal Studies at UCL Institute of Education.	
Variable code list	1	I Professional
	2	II Managerial and technical
	3	IIINM Skilled non-manual
	4	IIIM Skilled manual
	5	IV Partly skilled
	6	V Unskilled
	-666	Occup unclassifiable/ insuffic info/ armed forces/ carer/ unemplyed/ sick/ retired
	-777	No male head *(NCDS only)*



GAD-7 Anxiety

Over the last two weeks, how often have you been bothered by the following problems?	Not at all	Several days	More than half the days	Nearly every day
1. Feeling nervous, anxious, or on edge	0	1	2	3
2. Not being able to stop or control worrying	0	1	2	3
3. Worrying too much about different things	0	1	2	3
4. Trouble relaxing	0	1	2	3
5. Being so restless that it is hard to sit still	0	1	2	3
6. Becoming easily annoyed or irritable	0	1	2	3
7. Feeling afraid, as if something awful might happen	0	1	2	3



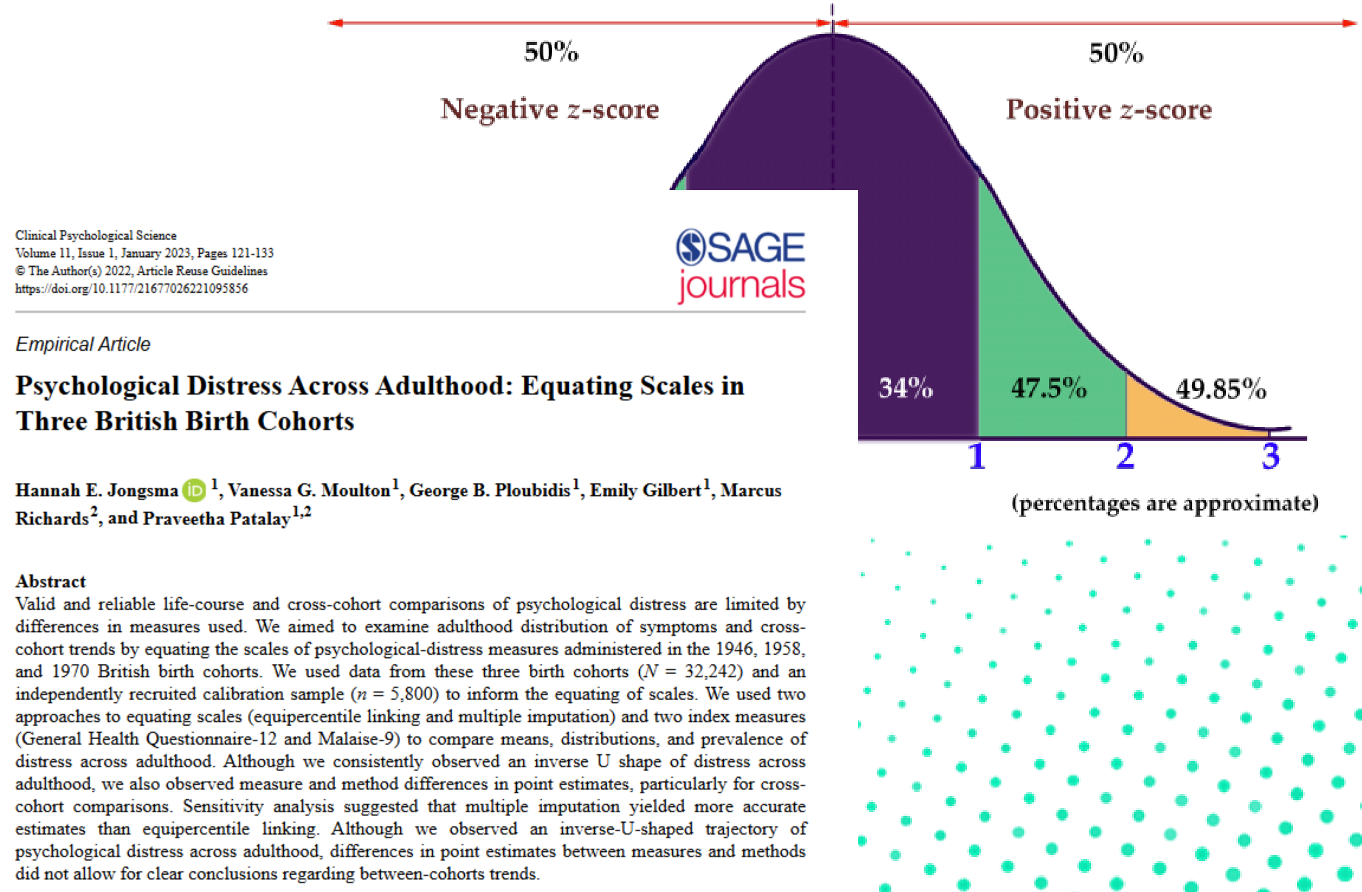
Beck Anxiety Inventory (BAI)

Below is a list of common symptoms of anxiety. Please carefully read each item in the list. Indicate how much you have been bothered by that symptom during the past month, including today, by circling the number in the corresponding space in the column next to each symptom.

	Not at all	Mildly, but it didn't bother me much	Moderately – it wasn't pleasant at times	Severely – it bothered me a lot
Numbness or tingling	0	1	2	3
Feeling hot	0	1	2	3
Wobbliness in legs	0	1	2	3
Unable to relax	0	1	2	3
Fear of worst happening	0	1	2	3
Dizzy or lightheaded	0	1	2	3
Heart pounding / racing	0	1	2	3
Unsteady	0	1	2	3
Terrified or afraid	0	1	2	3
Nervous	0	1	2	3
Feeling of choking	0	1	2	3
Hands trembling	0	1	2	3
Shaky / unsteady	0	1	2	3
Fear of losing control	0	1	2	3
Difficulty in breathing	0	1	2	3
Fear of dying	0	1	2	3
Scared	0	1	2	3
Indigestion	0	1	2	3

Harmonising scale data – different approaches

- Standardizing
- Scale Calibration
- Item-level retrospective harmonisation



Clinical Psychological Science
Volume 11, Issue 1, January 2023, Pages 121-133
© The Author(s) 2022, Article Reuse Guidelines
https://doi.org/10.1177/21677026221095856

SAGE journals

Empirical Article

Psychological Distress Across Adulthood: Equating Scales in Three British Birth Cohorts

Hannah E. Jongsma[1], Vanessa G. Moulton[1], George B. Ploubidis[1], Emily Gilbert[1], Marcus Richards[2], and Praveetha Patalay[1,2]

Abstract
Valid and reliable life-course and cross-cohort comparisons of psychological distress are limited by differences in measures used. We aimed to examine adulthood distribution of symptoms and cross-cohort trends by equating the scales of psychological-distress measures administered in the 1946, 1958, and 1970 British birth cohorts. We used data from these three birth cohorts (N = 32,242) and an independently recruited calibration sample (n = 5,800) to inform the equating of scales. We used two approaches to equating scales (equipercentile linking and multiple imputation) and two index measures (General Health Questionnaire-12 and Malaise-9) to compare means, distributions, and prevalence of distress across adulthood. Although we consistently observed an inverse U shape of distress across adulthood, we also observed measure and method differences in point estimates, particularly for cross-cohort comparisons. Sensitivity analysis suggested that multiple imputation yielded more accurate estimates than equipercentile linking. Although we observed an inverse-U-shaped trajectory of psychological distress across adulthood, differences in point estimates between measures and methods did not allow for clear conclusions regarding between-cohorts trends.

Harmonisation – types of data

Resource report

closer

The home of longitudinal research

Harmonisation and measurement properties of mental health measures in six British cohorts

Eoin McElroy[1], Aase Villadsen[1], Praveetha Patalay[1, 2], Alissa Goodman[1], Marcus Richards[2], Kate Northstone[3], Pasco Fearon[4], Marc Tibber[4], Dawid Gondek[1], George B. Ploubidis[1]

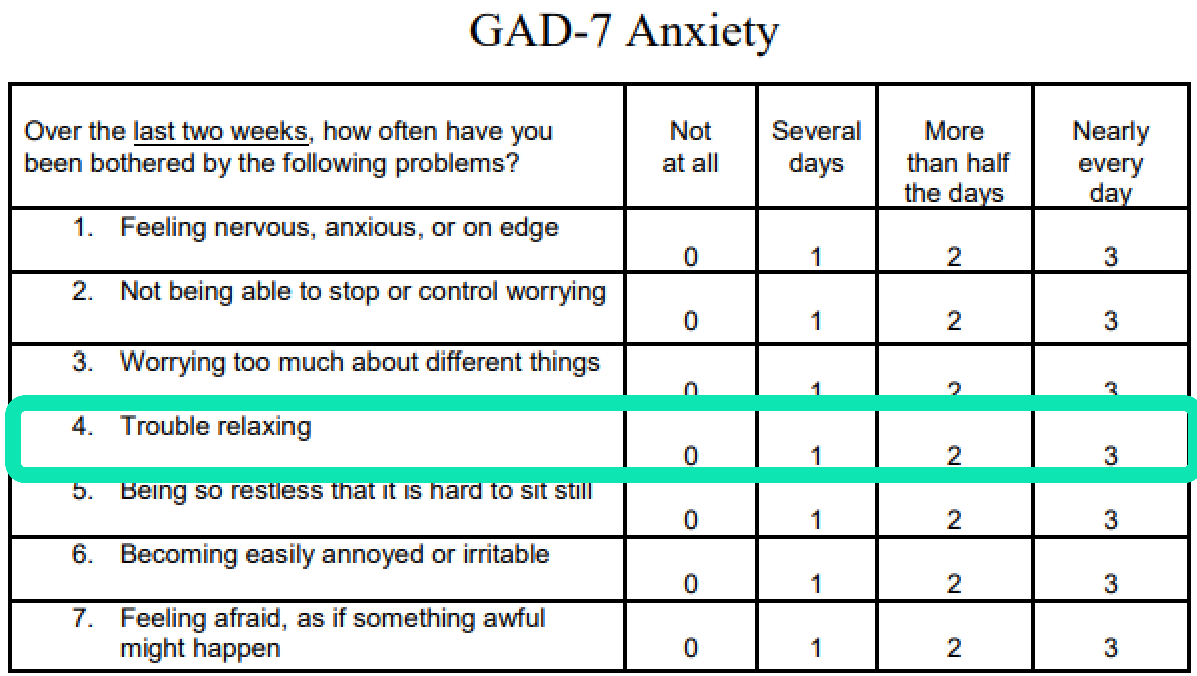
[1] Centre for Longitudinal Studies, University College London

[2] MRC Unit for Lifelong Health and Ageing, University College London

[3] MRC Integrative Epidemiology Unit, University of Bristol

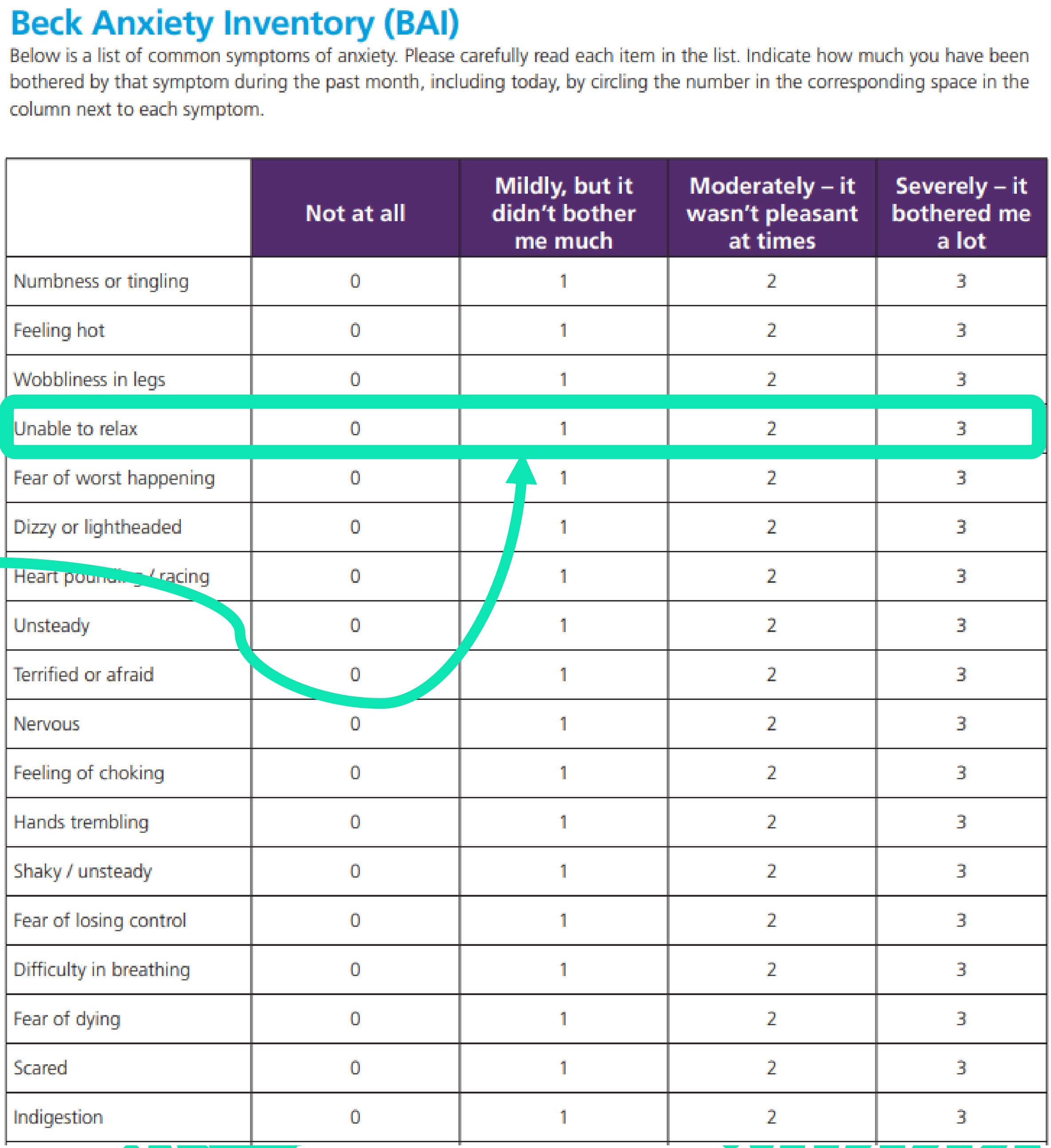
[4] Faculty of Brain Sciences, University College London

McElroy, E., Villadsen, A., Patalay, P., Goodman, A., Richards, M., Northstone, K., ... & Ploubidis, G. B. (2020). Harmonisation and measurement properties of mental health measures in six British cohorts. *UK: CLOSER*.



GAD-7 Anxiety

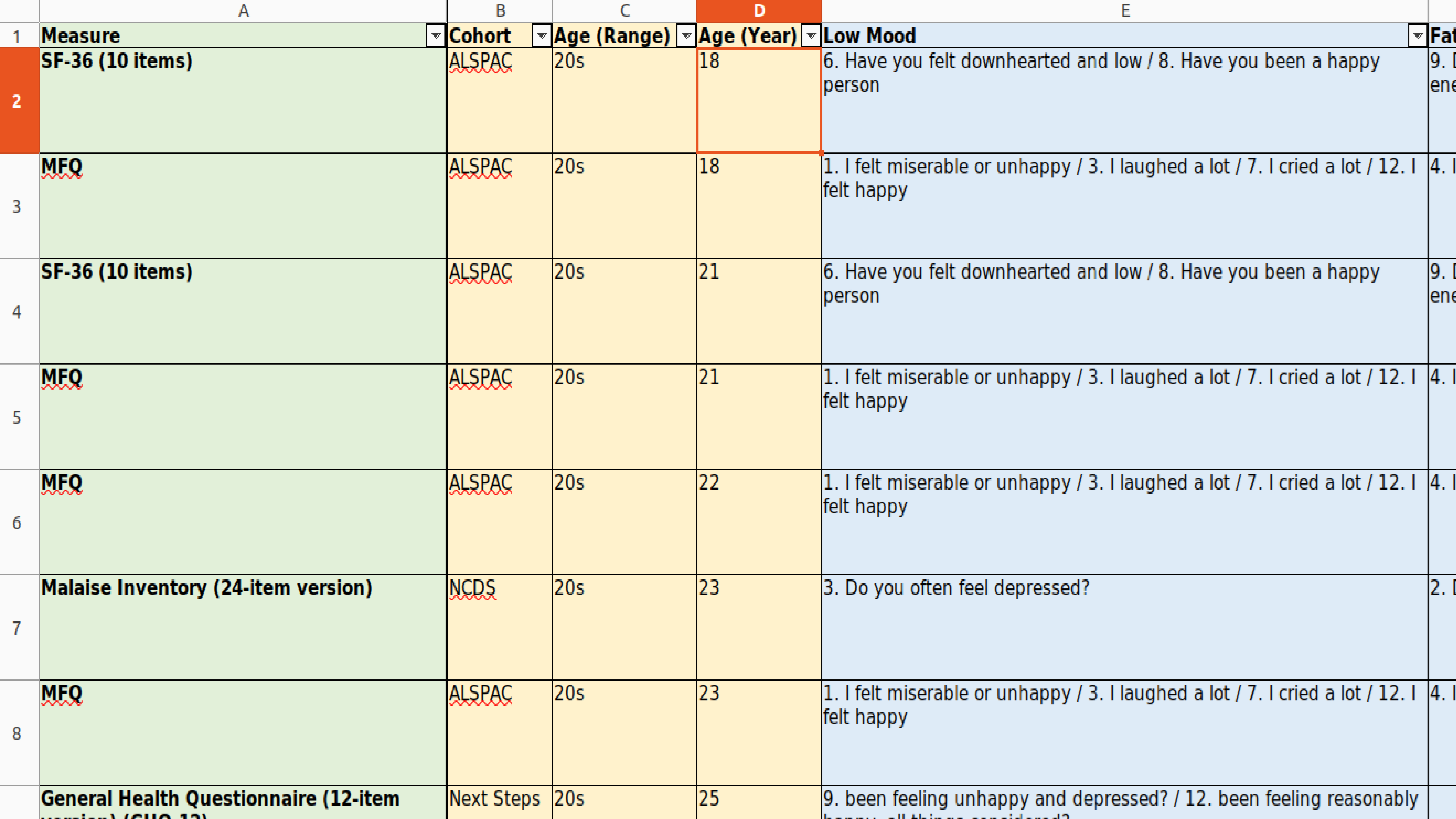
Over the last two weeks, how often have you been bothered by the following problems?	Not at all	Several days	More than half the days	Nearly every day
1. Feeling nervous, anxious, or on edge	0	1	2	3
2. Not being able to stop or control worrying	0	1	2	3
3. Worrying too much about different things	0	1	2	3
4. Trouble relaxing	0	1	2	3
5. Being so restless that it is hard to sit still	0	1	2	3
6. Becoming easily annoyed or irritable	0	1	2	3
7. Feeling afraid, as if something awful might happen	0	1	2	3



Beck Anxiety Inventory (BAI)

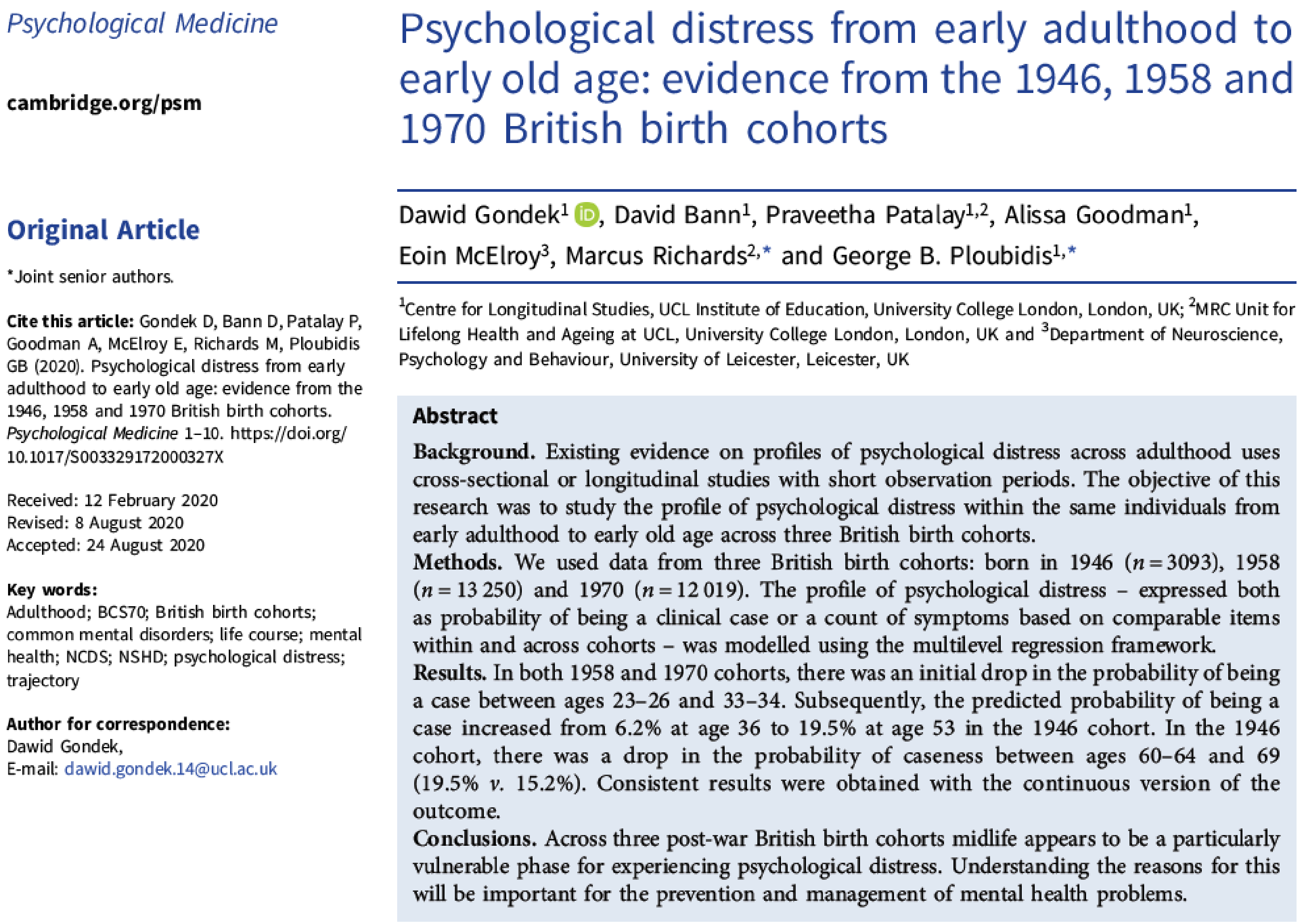
Below is a list of common symptoms of anxiety. Please carefully read each item in the list. Indicate how much you have been bothered by that symptom during the past month, including today, by circling the number in the corresponding space in the column next to each symptom.

	Not at all	Mildly, but it didn't bother me much	Moderately – it wasn't pleasant at times	Severely – it bothered me a lot
Numbness or tingling	0	1	2	3
Feeling hot	0	1	2	3
Wobbliness in legs	0	1	2	3
Unable to relax	0	1	2	3
Fear of worst happening	0	1	2	3
Dizzy or lightheaded	0	1	2	3
Heart pounding / racing	0	1	2	3
Unsteady	0	1	2	3
Terrified or afraid	0	1	2	3
Nervous	0	1	2	3
Feeling of choking	0	1	2	3
Hands trembling	0	1	2	3
Shaky / unsteady	0	1	2	3
Fear of losing control	0	1	2	3
Difficulty in breathing	0	1	2	3
Fear of dying	0	1	2	3
Scared	0	1	2	3
Indigestion	0	1	2	3



	A	B	C	D	E	
1	**Measure**	**Cohort**	**Age (Range)**	**Age (Year)**	**Low Mood**	**Fat**
2	**SF-36 (10 items)**	ALSPAC	20s	18	6. Have you felt downhearted and low / 8. Have you been a happy person	9. I ene
3	**MFQ**	ALSPAC	20s	18	1. I felt miserable or unhappy / 3. I laughed a lot / 7. I cried a lot / 12. I felt happy	4. I
4	**SF-36 (10 items)**	ALSPAC	20s	21	6. Have you felt downhearted and low / 8. Have you been a happy person	9. I ene
5	**MFQ**	ALSPAC	20s	21	1. I felt miserable or unhappy / 3. I laughed a lot / 7. I cried a lot / 12. I felt happy	4. I
6	**MFQ**	ALSPAC	20s	22	1. I felt miserable or unhappy / 3. I laughed a lot / 7. I cried a lot / 12. I felt happy	4. I
7	**Malaise Inventory (24-item version)**	NCDS	20s	23	3. Do you often feel depressed?	2. I
8	**MFQ**	ALSPAC	20s	23	1. I felt miserable or unhappy / 3. I laughed a lot / 7. I cried a lot / 12. I felt happy	4. I
	General Health Questionnaire (12-item version) (GHQ-12)	Next Steps	20s	25	9. been feeling unhappy and depressed? / 12. been feeling reasonably happy, all things considered?	

Case study



Psychological Medicine

cambridge.org/psm

Original Article

*Joint senior authors.

Cite this article: Gondek D, Bann D, Patalay P, Goodman A, McElroy E, Richards M, Ploubidis GB (2020). Psychological distress from early adulthood to early old age: evidence from the 1946, 1958 and 1970 British birth cohorts. *Psychological Medicine* 1–10. https://doi.org/10.1017/S003329172000327X

Received: 12 February 2020
Revised: 8 August 2020
Accepted: 24 August 2020

Key words:
Adulthood; BCS70; British birth cohorts; common mental disorders; life course; mental health; NCDS; NSHD; psychological distress; trajectory

Author for correspondence:
Dawid Gondek,
E-mail: dawid.gondek.14@ucl.ac.uk

Psychological distress from early adulthood to early old age: evidence from the 1946, 1958 and 1970 British birth cohorts

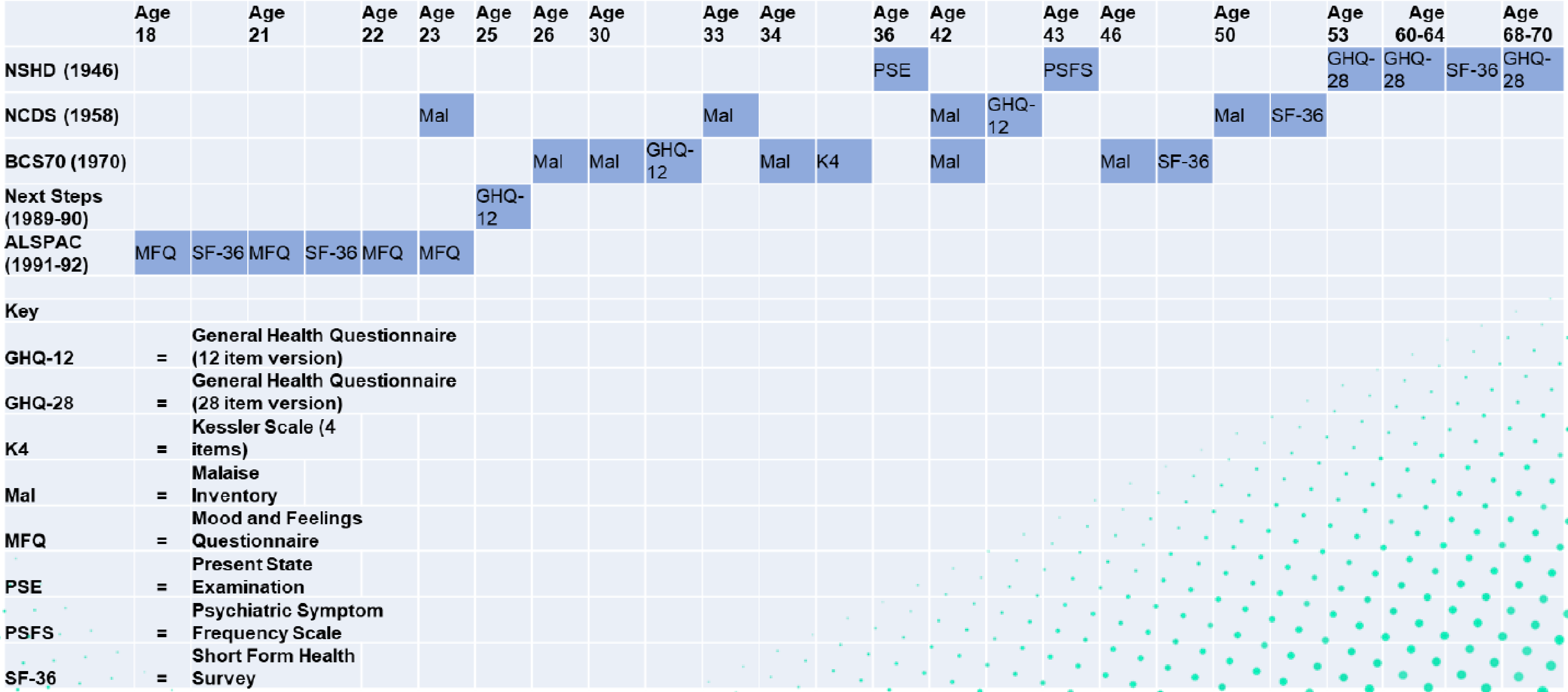
Dawid Gondek[1], David Bann[1], Praveetha Patalay[1,2], Alissa Goodman[1], Eoin McElroy[3], Marcus Richards[2,*] and George B. Ploubidis[1,*]

[1]Centre for Longitudinal Studies, UCL Institute of Education, University College London, London, UK; [2]MRC Unit for Lifelong Health and Ageing at UCL, University College London, London, UK and [3]Department of Neuroscience, Psychology and Behaviour, University of Leicester, Leicester, UK

Abstract

Background. Existing evidence on profiles of psychological distress across adulthood uses cross-sectional or longitudinal studies with short observation periods. The objective of this research was to study the profile of psychological distress within the same individuals from early adulthood to early old age across three British birth cohorts.
Methods. We used data from three British birth cohorts: born in 1946 ($n = 3093$), 1958 ($n = 13\,250$) and 1970 ($n = 12\,019$). The profile of psychological distress – expressed both as probability of being a clinical case or a count of symptoms based on comparable items within and across cohorts – was modelled using the multilevel regression framework.
Results. In both 1958 and 1970 cohorts, there was an initial drop in the probability of being a case between ages 23–26 and 33–34. Subsequently, the predicted probability of being a case increased from 6.2% at age 36 to 19.5% at age 53 in the 1946 cohort. In the 1946 cohort, there was a drop in the probability of caseness between ages 60–64 and 69 (19.5% *v.* 15.2%). Consistent results were obtained with the continuous version of the outcome.
Conclusions. Across three post-war British birth cohorts midlife appears to be a particularly vulnerable phase for experiencing psychological distress. Understanding the reasons for this will be important for the prevention and management of mental health problems.

Case study



	Age 18		Age 21		Age 22	Age 23	Age 25	Age 26	Age 30		Age 33	Age 34		Age 36	Age 42		Age 43	Age 46		Age 50		Age 53	Age 60-64		Age 68-70
NSHD (1946)														PSE			PSFS					GHQ-28	GHQ-28	SF-36	GHQ-28
NCDS (1958)						Mal					Mal				Mal	GHQ-12				Mal	SF-36				
BCS70 (1970)								Mal	Mal	GHQ-12		Mal	K4		Mal			Mal	SF-36						
Next Steps (1989-90)							GHQ-12																		
ALSPAC (1991-92)	MFQ	SF-36	MFQ	SF-36	MFQ	MFQ																			
Key																									
GHQ-12	=	General Health Questionnaire (12 item version)																							
GHQ-28	=	General Health Questionnaire (28 item version)																							
K4	=	Kessler Scale (4 items)																							
Mal	=	Malaise Inventory																							
MFQ	=	Mood and Feelings Questionnaire																							
PSE	=	Present State Examination																							
PSFS	=	Psychiatric Symptom Frequency Scale																							
SF-36	=	Short Form Health Survey																							

Figure 2. Overview of mental health measures administered throughout adulthood in six British cohort studies (all measures are self-reports)

(McElroy et al., 2020)

Case study

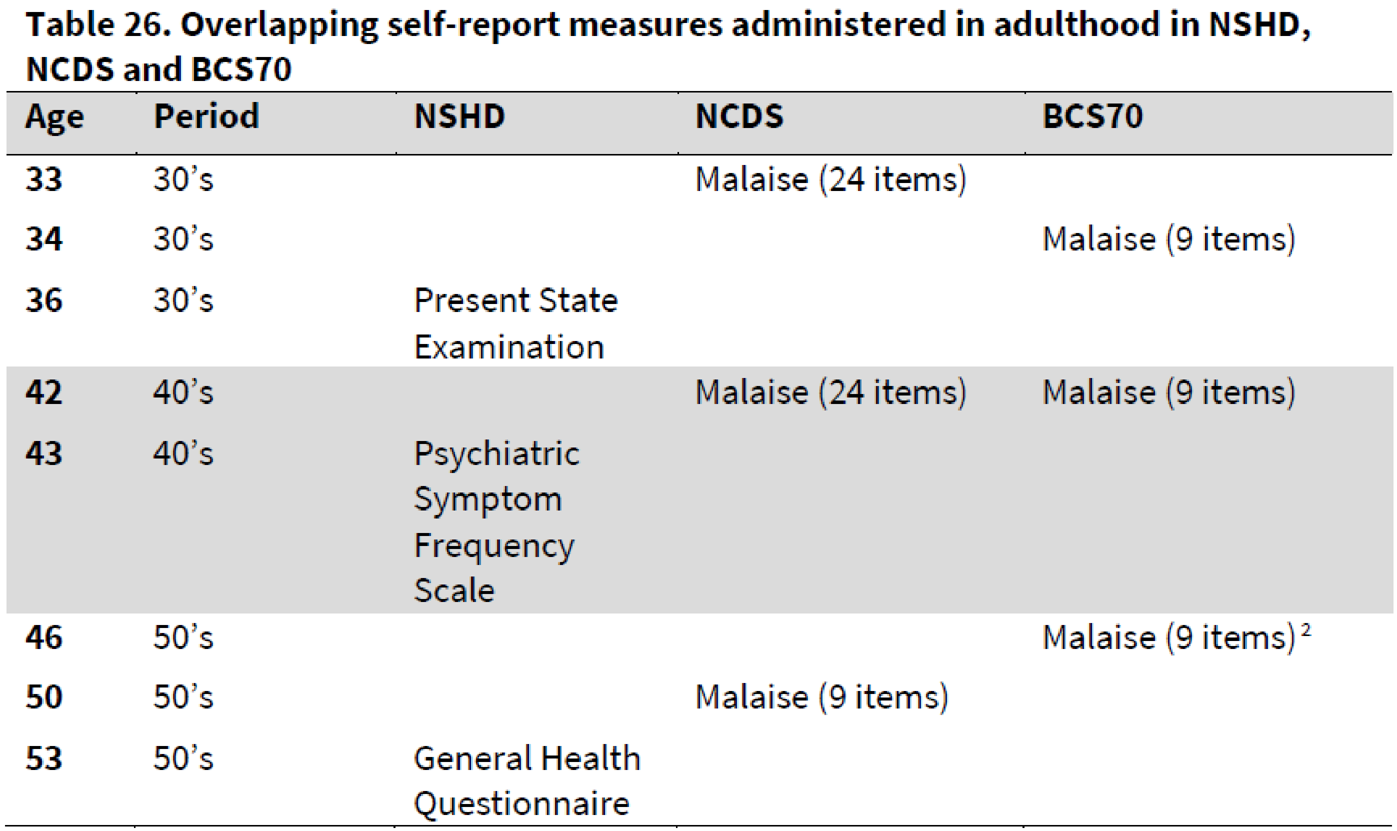


Table 26. Overlapping self-report measures administered in adulthood in NSHD, NCDS and BCS70

Age	Period	NSHD	NCDS	BCS70
33	30’s		Malaise (24 items)	
34	30’s			Malaise (9 items)
36	30’s	Present State Examination		
42	40’s		Malaise (24 items)	Malaise (9 items)
43	40’s	Psychiatric Symptom Frequency Scale		
46	50’s			Malaise (9 items)[2]
50	50’s		Malaise (9 items)	
53	50’s	General Health Questionnaire		

(McElroy et al., 2020)

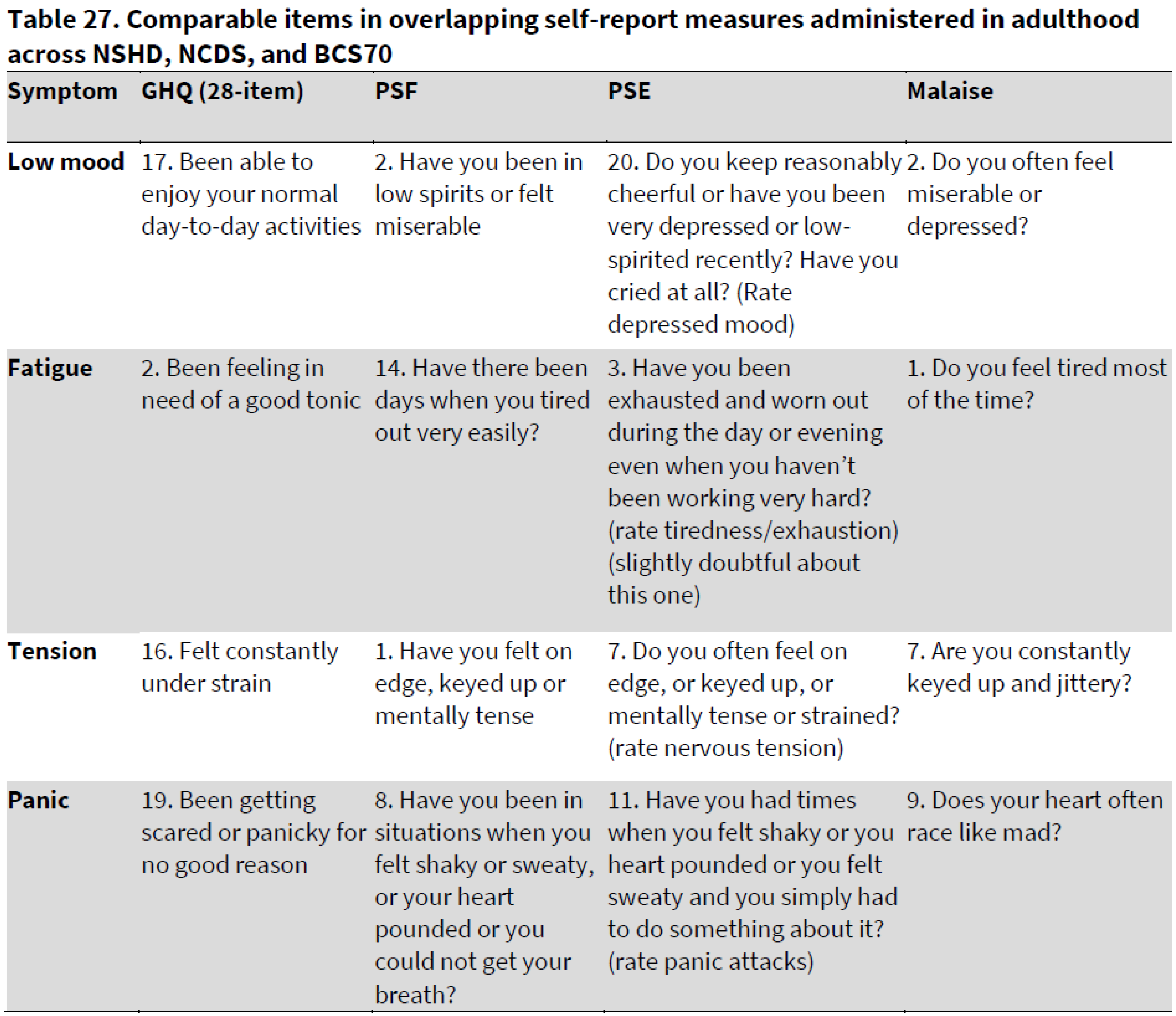


Table 27. Comparable items in overlapping self-report measures administered in adulthood across NSHD, NCDS, and BCS70

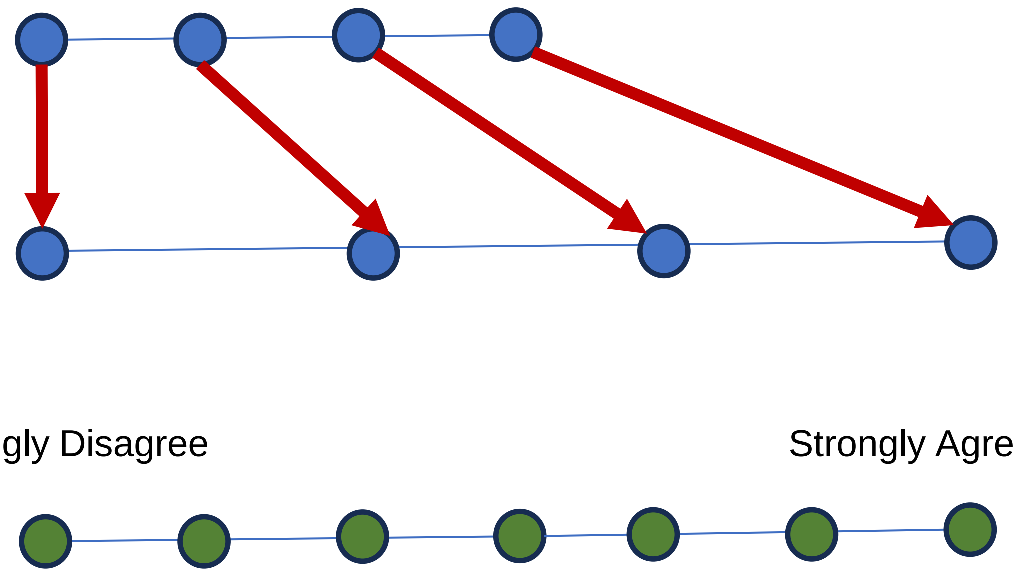
Symptom	GHQ (28-item)	PSF	PSE	Malaise
Low mood	17. Been able to enjoy your normal day-to-day activities	2. Have you been in low spirits or felt miserable	20. Do you keep reasonably cheerful or have you been very depressed or low-spirited recently? Have you cried at all? (Rate depressed mood)	2. Do you often feel miserable or depressed?
Fatigue	2. Been feeling in need of a good tonic	14. Have there been days when you tired out very easily?	3. Have you been exhausted and worn out during the day or evening even when you haven't been working very hard? (rate tiredness/exhaustion) (slightly doubtful about this one)	1. Do you feel tired most of the time?
Tension	16. Felt constantly under strain	1. Have you felt on edge, keyed up or mentally tense	7. Do you often feel on edge, or keyed up, or mentally tense or strained? (rate nervous tension)	7. Are you constantly keyed up and jittery?
Panic	19. Been getting scared or panicky for no good reason	8. Have you been in situations when you felt shaky or sweaty, or your heart pounded or you could not get your breath?	11. Have you had times when you felt shaky or you heart pounded or you felt sweaty and you simply had to do something about it? (rate panic attacks)	9. Does your heart often race like mad?

(McElroy et al., 2020)

Harmonisation of response options

Scale / Example Item	Original Response	Recoding	Harmonised response	Recoding	Original Response	Scale / Example Item
Psychiatric Symptom Frequency Scale "2. Have you been in low spirits or felt miserable"	0 = never in the last year. 1 = up to 10 days in total, less than once a month. 2 = a spell up to one month, once or twice a month, 'a months worth'. 3 = a spell up to four months, once or twice a week, three to ten times a month. 4 = a spell of over four months, three or more times a week, 11 or more times a month. 5 = every day in the last year.	0 = 0 1 to 5 = 1	0 = Absence 1 = Presence	NA	0 =No 1 = Yes	Malaise "2. Do you often feel iserable or depressed"

Harmonisation of response options



Disagree

Agree agree

Strongly Disagree

Strongly Agree

$$\textbf{stretch}(x) = \frac{x\mathrm{K}y}{\mathrm{K}x}$$

(Singh, 2022)

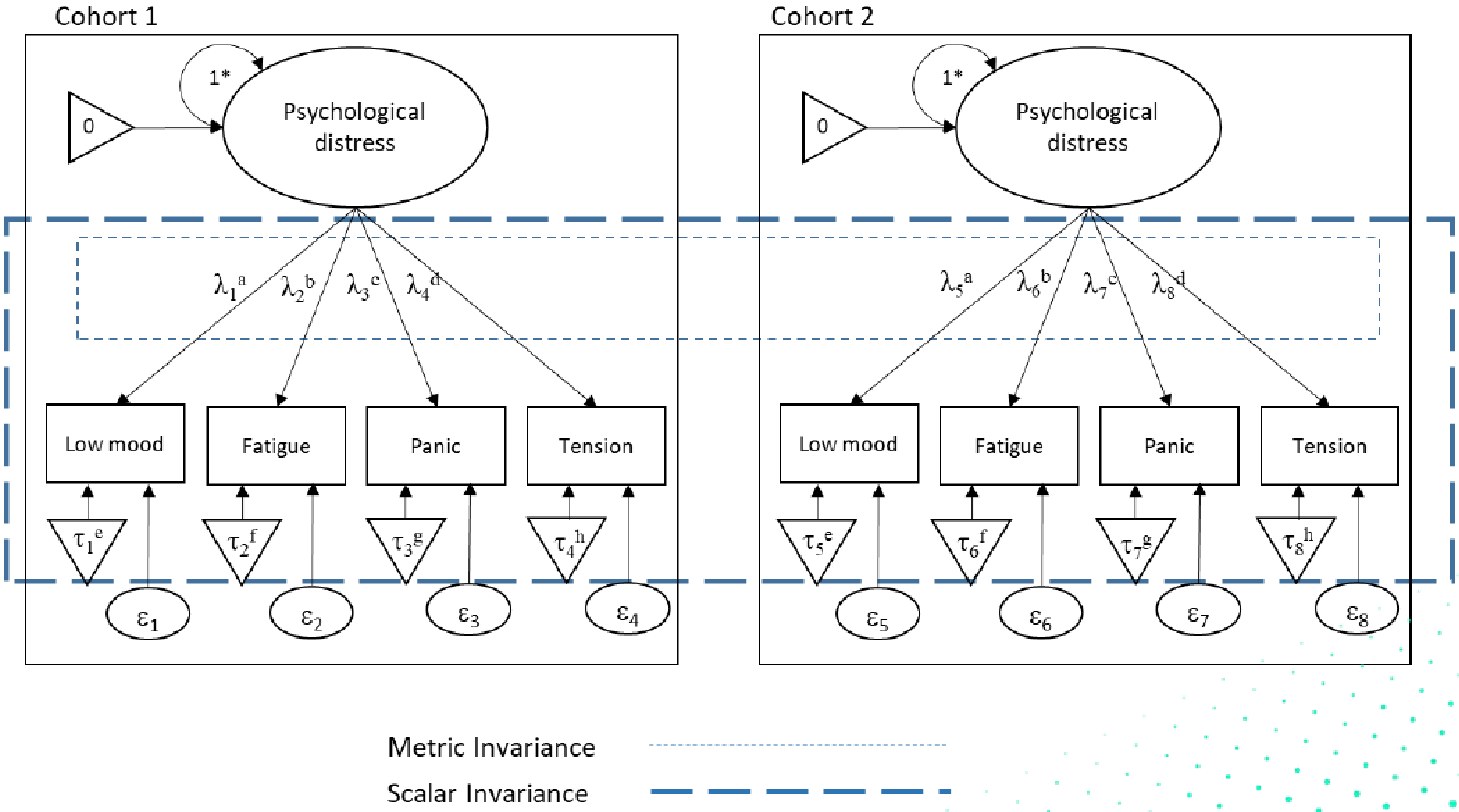


Figure 4. Graphical illustration of multiple group confirmatory factor analysis, with four measured indicators of a general psychological distress factor, assessed across two cohorts

λ = Factor loadings; τ = Thresholds; ⊠ = residuals (theta parameterisation); a-d = loadings held equal across cohorts in test for metric invariance; e-h = thresholds held equal across cohorts in test for scalar invariance

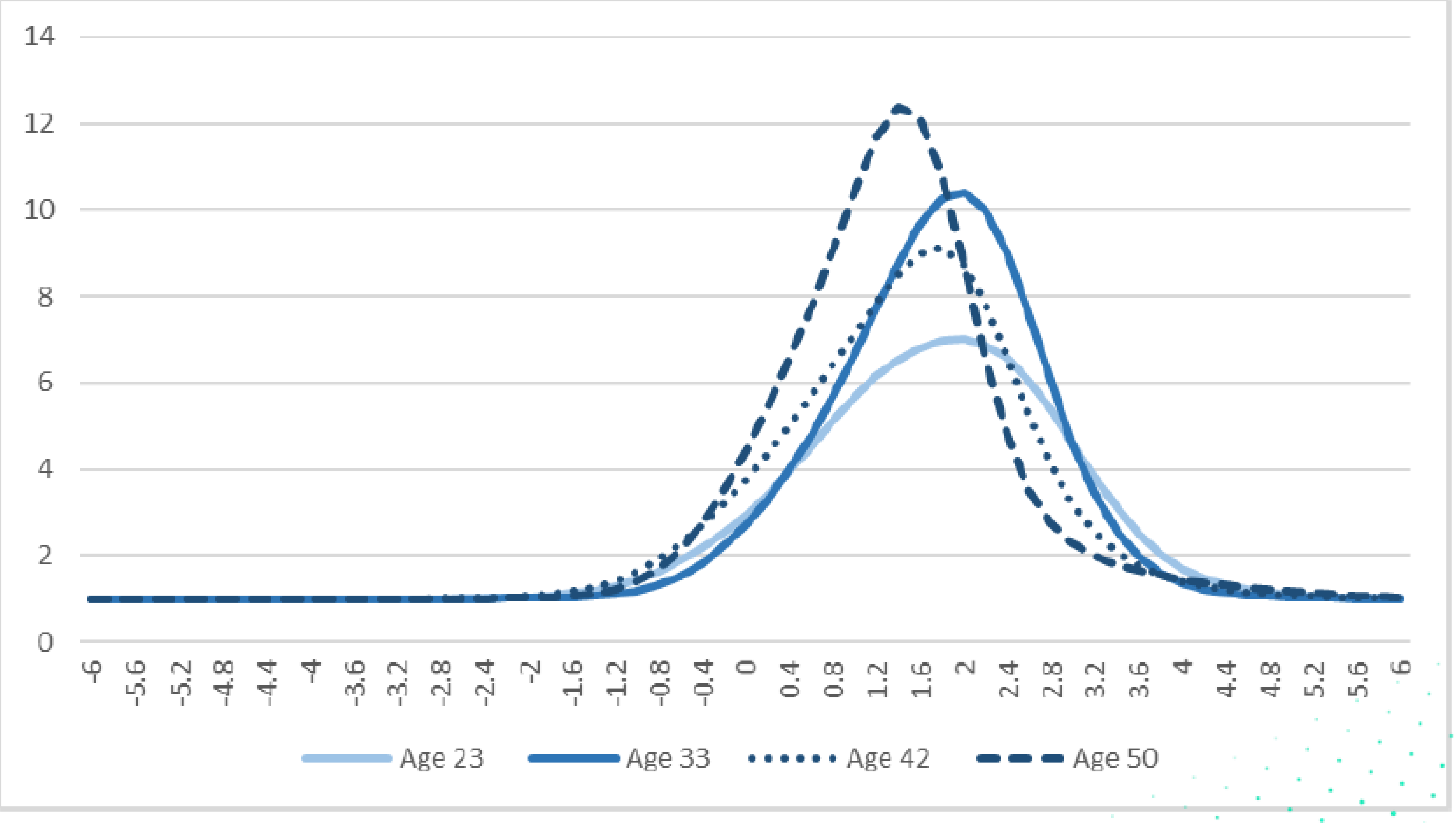
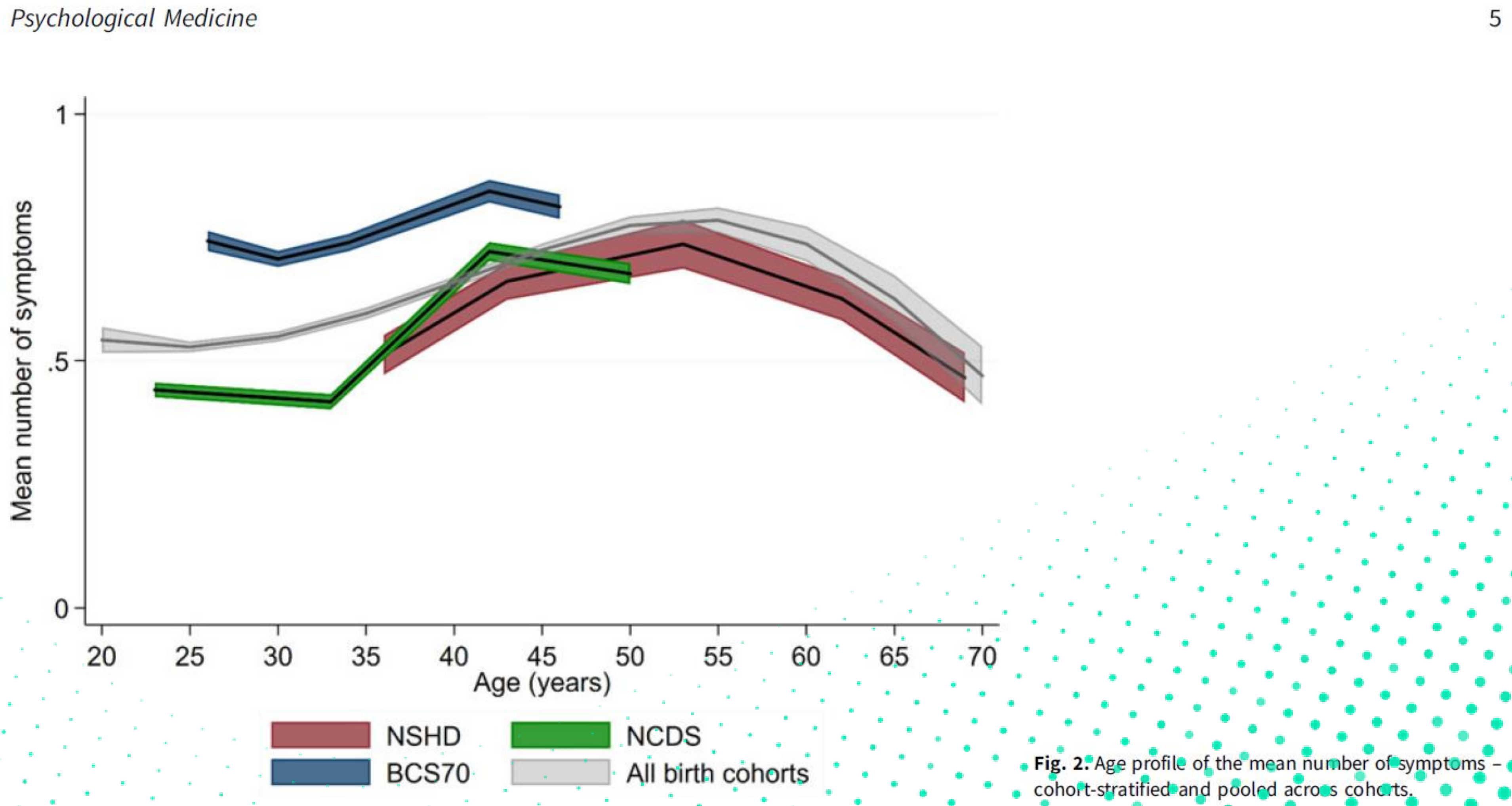


Figure 24. TIFs for Malaise Inventory (9-item version) in NCDS

Case study



Psychological Medicine 5

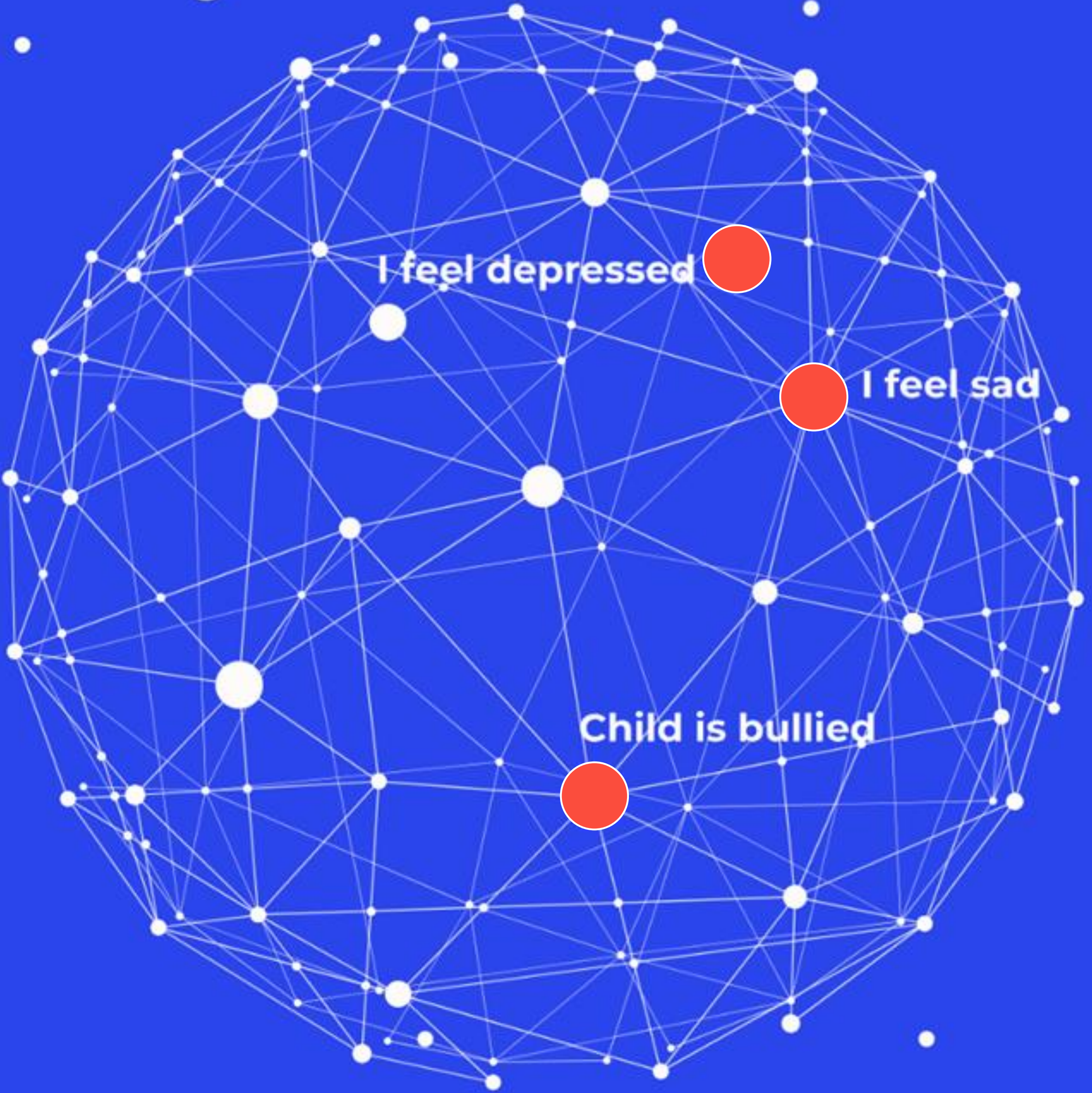
Fig. 2. Age profile of the mean number of symptoms – cohort-stratified and pooled across cohorts.

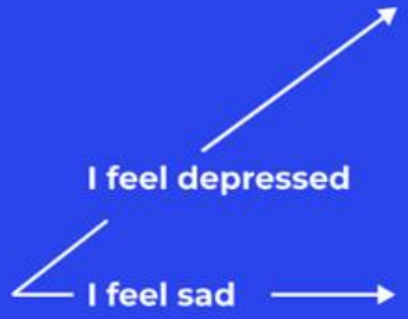
HARMONY

Harmony

Open source NLP/AI tool for psychologists and social and health sciences
Data discovery and harmonisation

harmonydata.ac.uk
github.com/harmonydata/

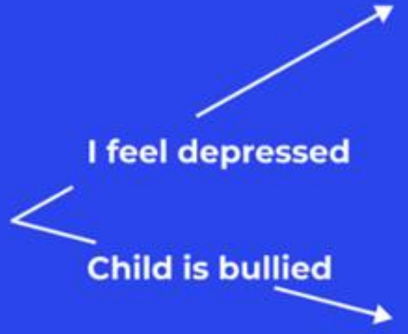




Sentences have a related meaning

Vectors point in a similar direction

Cosine score is high – close to 1



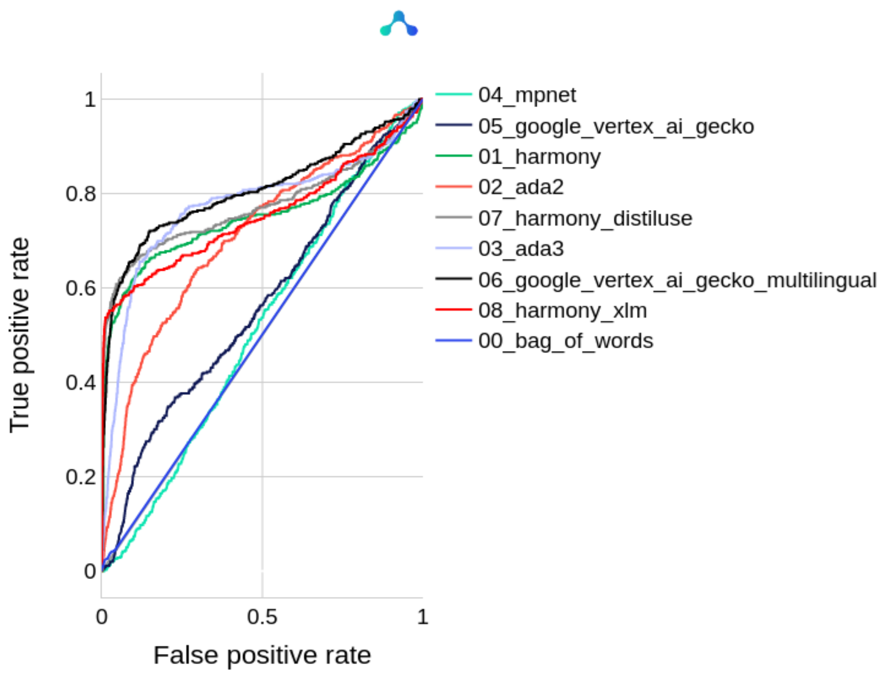
Sentences have unrelated meaning

Vectors are orthogonal

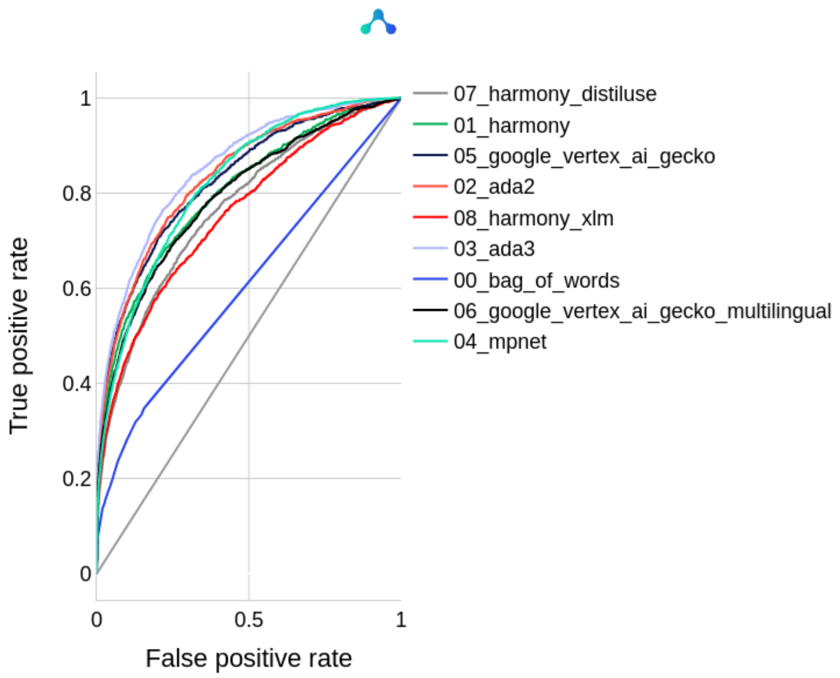
Cosine score is low – closer to 0

Evaluating Harmony

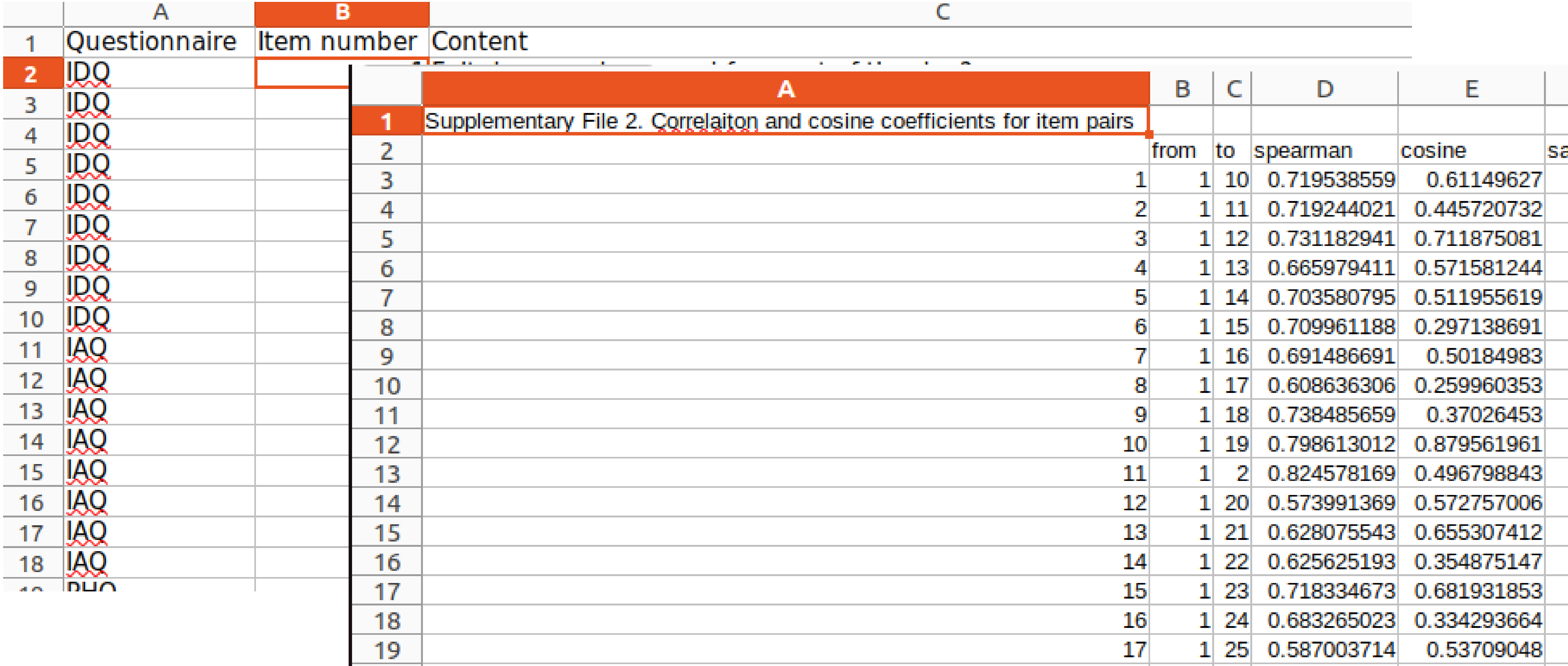
ROC on GAD 7 multilingual dataset



ROC on McElroy et al childhood dataset



Real correlations



	A	B	C
1	Questionnaire	Item number	Content
2	IDQ		
3	IDQ		
4	IDQ		
5	IDQ		
6	IDQ		
7	IDQ		
8	IDQ		
9	IDQ		
10	IDQ		
11	IAQ		
12	IAQ		
13	IAQ		
14	IAQ		
15	IAQ		
16	IAQ		
17	IAQ		
18	IAQ		

	A	B	C	D	E
1	Supplementary File 2. Correlaiton and cosine coefficients for item pairs				
2		from	to	spearman	cosine
3	1	1	10	0.719538559	0.61149627
4	2	1	11	0.719244021	0.445720732
5	3	1	12	0.731182941	0.711875081
6	4	1	13	0.665979411	0.571581244
7	5	1	14	0.703580795	0.511955619
8	6	1	15	0.709961188	0.297138691
9	7	1	16	0.691486691	0.50184983
10	8	1	17	0.608636306	0.259960353
11	9	1	18	0.738485659	0.37026453
12	10	1	19	0.798613012	0.879561961
13	11	1	2	0.824578169	0.496798843
14	12	1	20	0.573991369	0.572757006
15	13	1	21	0.628075543	0.655307412
16	14	1	22	0.625625193	0.354875147
17	15	1	23	0.718334673	0.681931853
18	16	1	24	0.683265023	0.334293664
19	17	1	25	0.587003714	0.53709048

Real correlations

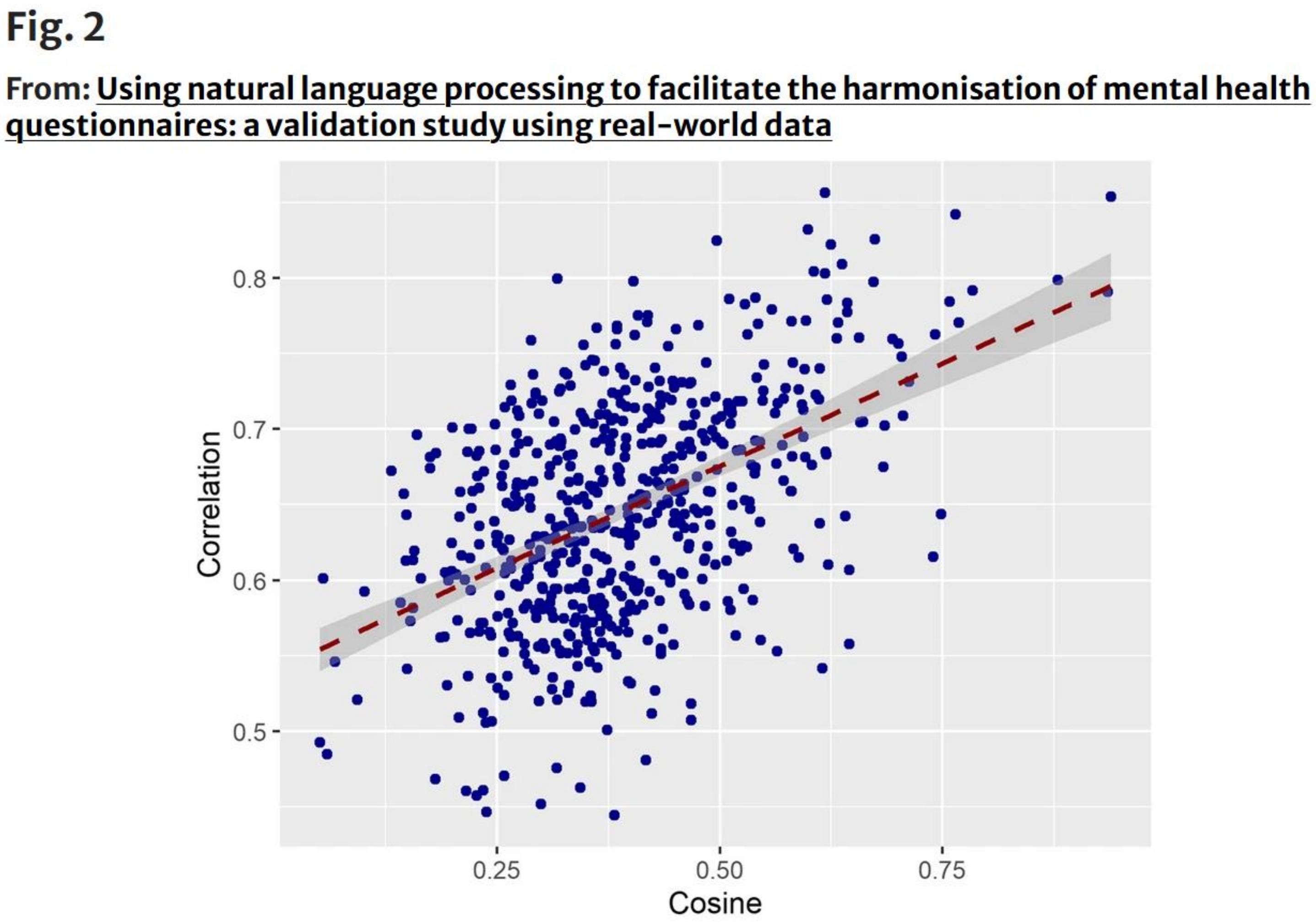


Fig. 2

From: Using natural language processing to facilitate the harmonisation of mental health questionnaires: a validation study using real-world data

McElroy, E., Wood, T., Bond, R., Mulvenna, M., Shevlin, M., Ploubidis, G. B., ... & Moltrecht, B. (2024). Using natural language processing to facilitate the harmonisation of mental health questionnaires: a validation study using real-world data. *BMC psychiatry*, *24*(1), 530.

Open source

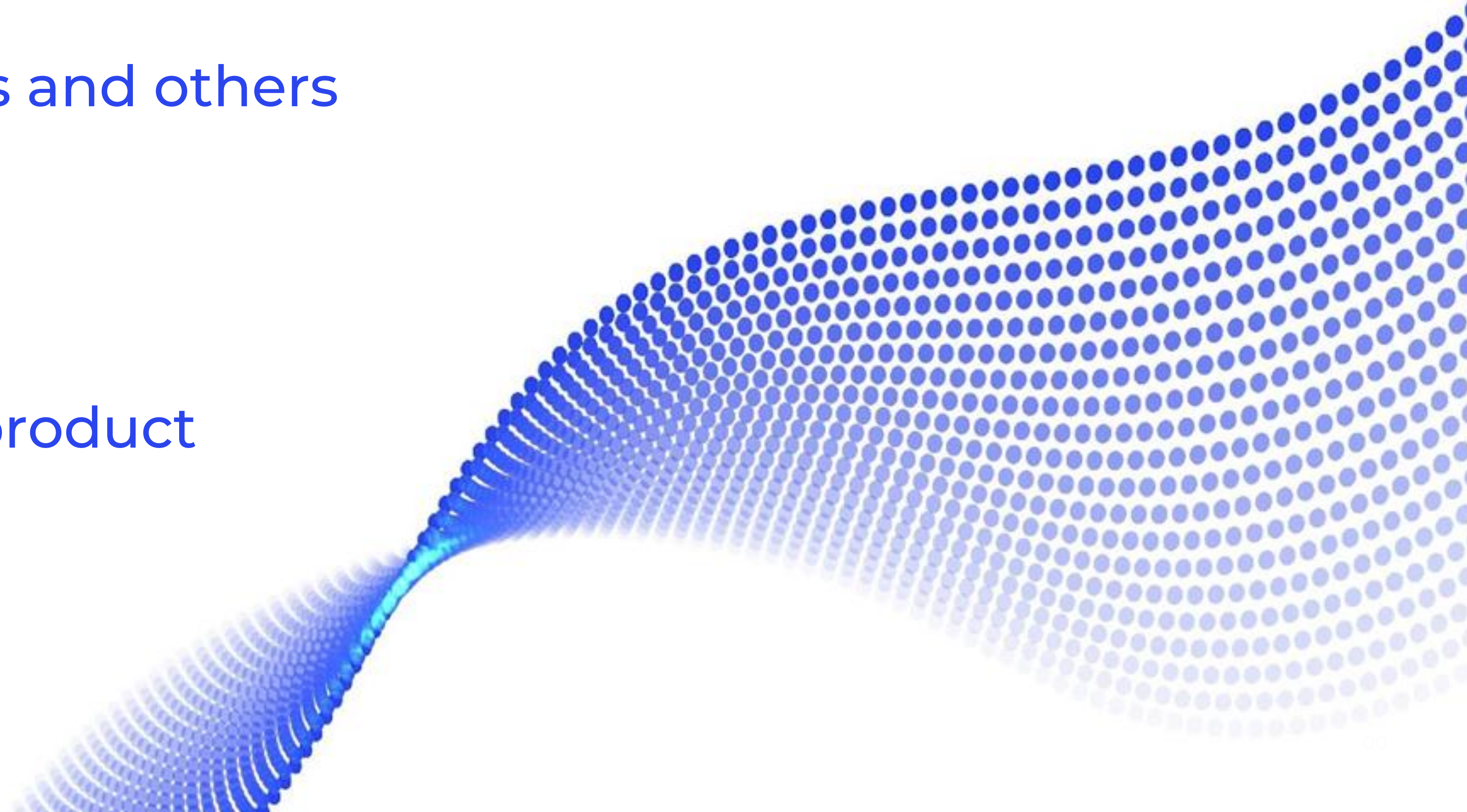
https://github.com/harmonydata

Free for psychologists and others around the world

MIT License

It’s not a monetised product

Hackathons



Integrations

Currently pulling data from

- UKDS
- HDR UK
- ADR UK
- Mental Health Catalogue
- UKLLC
- Closer

Future integrations

- Australian longitudinal studies on anxiety (from University of Sydney)

Possibly other sources such as

- Dementias Platform UK Cohort Directory
- Institute for Fiscal Studies
- UKRI Cohort Directory

UK Data Service

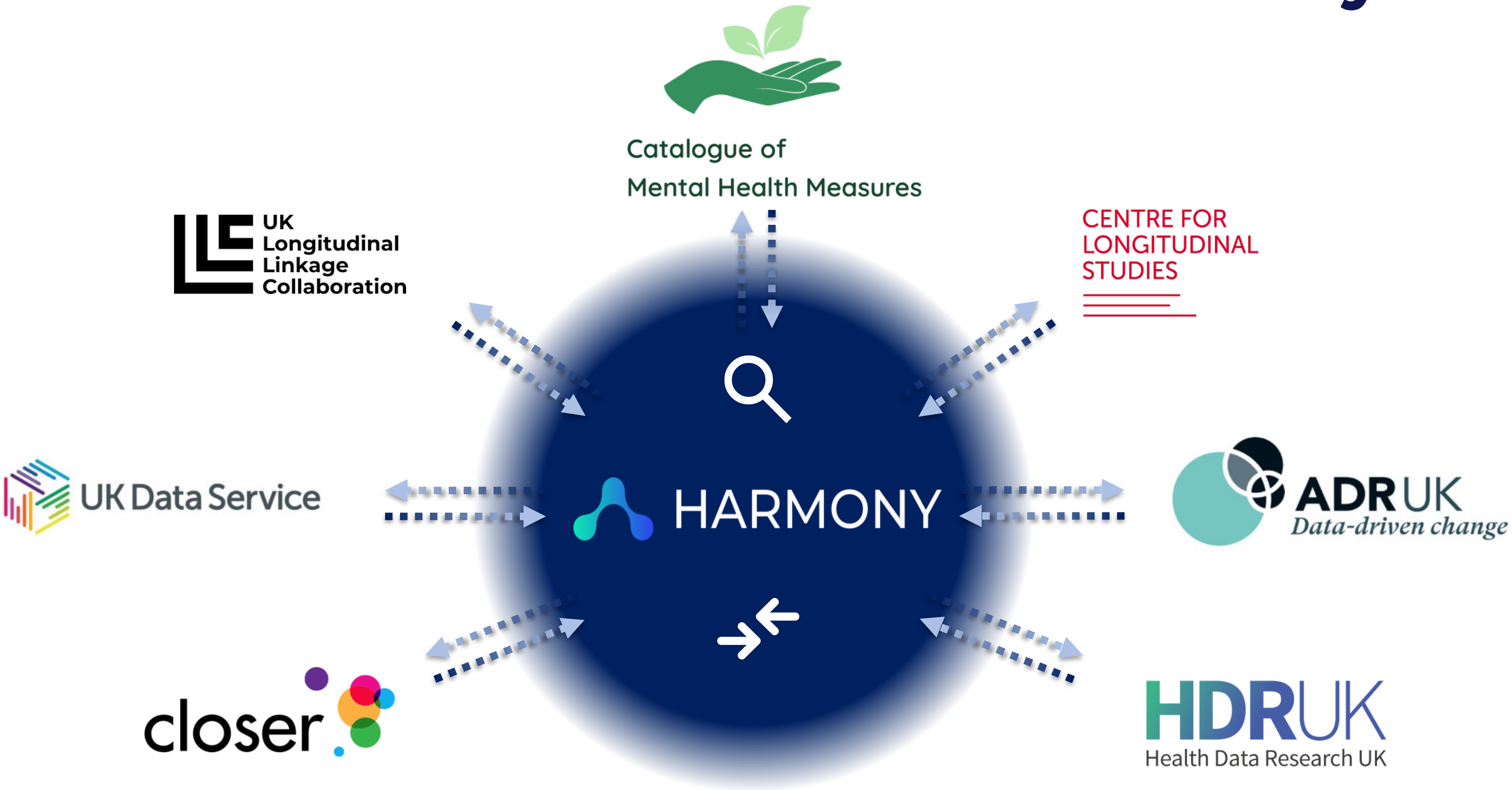
HDRUK
Health Data Research UK

ADR UK

UK
Longitudinal
Linkage
Collaboration

closer
The home of
longitudinal research

Next Steps: Building Bi-directional pathways for better harmonisation and discovery



HARMONY

Co-Design Harmony

Join our co-design process,
sign up to our mailing list at:

https://harmonydata.ac.uk

UKRI Economic and Social Research Council

wellcome

Ulster University

CENTRE FOR LONGITUDINAL STUDIES

UK Longitudinal Linkage Collaboration

Fast Data Science

Uses outside longitudinal studies?

- Market research? (Surveys about new products?)
- Pharma?
  - Informed Consent Forms?
  - Inclusion criteria?
  - Endpoints?
- Finance?
- Legal?
- National Archives?
- Your industry?

HARMONY

Thank you for listening!

harmonydata.ac.uk

discord.gg/harmonydata

linkedin.com/company/harmonydata

github.com/harmonydata/harmony

@harmony_data

@harmony_data

e.mcelroy@ulster.ac.uk thomas@fastdatascience.com

References

Tomescu-Dubrow, I., Wolf, C., Slomczynski, K. M., & Jenkins, J. C. (Eds.). (2023). Survey Data Harmonization in the Social Sciences. John Wiley & Sons.

Fortier, I., Raina, P., Van den Heuvel, E. R., Griffith, L. E., Craig, C., Saliba, M., ... & Burton, P. (2017). Maelstrom Research guidelines for rigorous retrospective data harmonization. International journal of epidemiology, 46(1), 103-105.

Cohort and Longitudinal Studies Enhancement Resources. (2017). Harmonised Height, Weight and BMI in Five Longitudinal Cohort Studies: National Child Development Study, 1970 British Cohort Study and Millennium Cohort Study. [data collection]. UK Data Service. SN: 8207, DOI: http://doi.org/10.5255/UKDA-SN-8207-1

Singh, R. K. (2022). Harmonizing Single-Question Instruments for Latent Constructs With Equating Using Political Interest as an Example. In Survey Research Methods (Vol. 16, No. 3, pp. 353-369).

Bann, D., Johnson, W., Li, L., Kuh, D., & Hardy, R. (2018). Socioeconomic inequalities in childhood and adolescent body-mass index, weight, and height from 1953 to 2015: an analysis of four longitudinal, observational, British birth cohort studies. The Lancet Public Health, 3(4), e194-e203.

Cohort and Longitudinal Studies Enhancement Resources. (2022). CLOSER. [data series]. 8th Release. UK Data Service. SN: 2000111, DOI: http://doi.org/10.5255/UKDA-Series-2000111

Jongsma, H. E., Moulton, V. G., Ploubidis, G. B., Gilbert, E., Richards, M., & Patalay, P. (2023). Psychological Distress Across Adulthood: Equating Scales in Three British Birth Cohorts. Clinical Psychological Science, 11(1), 121-133.

McElroy, E., Villadsen, A., Patalay, P., Goodman, A., Richards, M., Northstone, K., ... & Ploubidis, G. B. (2020). Harmonisation and measurement properties of mental health measures in six British cohorts. UK: CLOSER.

Gondek, D., Bann, D., Patalay, P., Goodman, A., McElroy, E., Richards, M., & Ploubidis, G. B. (2022). Psychological distress from early adulthood to early old age: evidence from the 1946, 1958 and 1970 British birth cohorts. Psychological Medicine, 52(8), 1471-1480.