When psychometrics meets epidemiology, and the studies which result!

Tom Booth

Tom.booth@ed.ac.uk

Lecturer in Quantitative Research Methods,

Department of Psychology, University of Edinburgh

Who am I?

- MSc and PhD in Organisational Psychology – ESRC AQM Scholarship
  - Manchester Business School, University of Manchester
  - *Topic: Personality psychometrics*

- Post-doctoral Researcher
  - Centre for Cognitive Ageing and Cognitive Epidemiology, Department of Psychology, University of Edinburgh.
  - *Primary Research Topic: Cognitive ageing, brain imaging*.

- Lecturer Quantitative Research Methods
  - Department of Psychology, University of Edinburgh
  - *Primary Research Topics: Individual differences and health; cognitive ability and brain imaging; Psychometric methods and assessment.*

Journey of a talk….

- Psychometrics:
    - Performance of likert-type response scales for personality data.
    - Murray, Booth & Molenaar (2015)

- Epidemiology:
    - Allostatic load
        - Measurement: Booth, Starr & Deary (2013); (*Unpublished*)
        - Applications: Early life adversity (*Unpublished*)
        - Further applications

Journey of a talk….

- Methodological interlude:
  - The issue of optimal time scales.
- Individual differences and health:
  - Personality and Physical Health (review: Murray & Booth, 2015)
  - Personality, health behaviours and brain integrity (Booth, Mottus et al., 2014)
- Looking forward

Psychometrics

My spiritual home…

Middle response options

Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strong Disagree
1	2	3	4	5

Strongly Agree	Agree	Unsure	Disagree	Strong Disagree
1	2	3	4	5

Yes	?	No
1	2	3

Middle response options

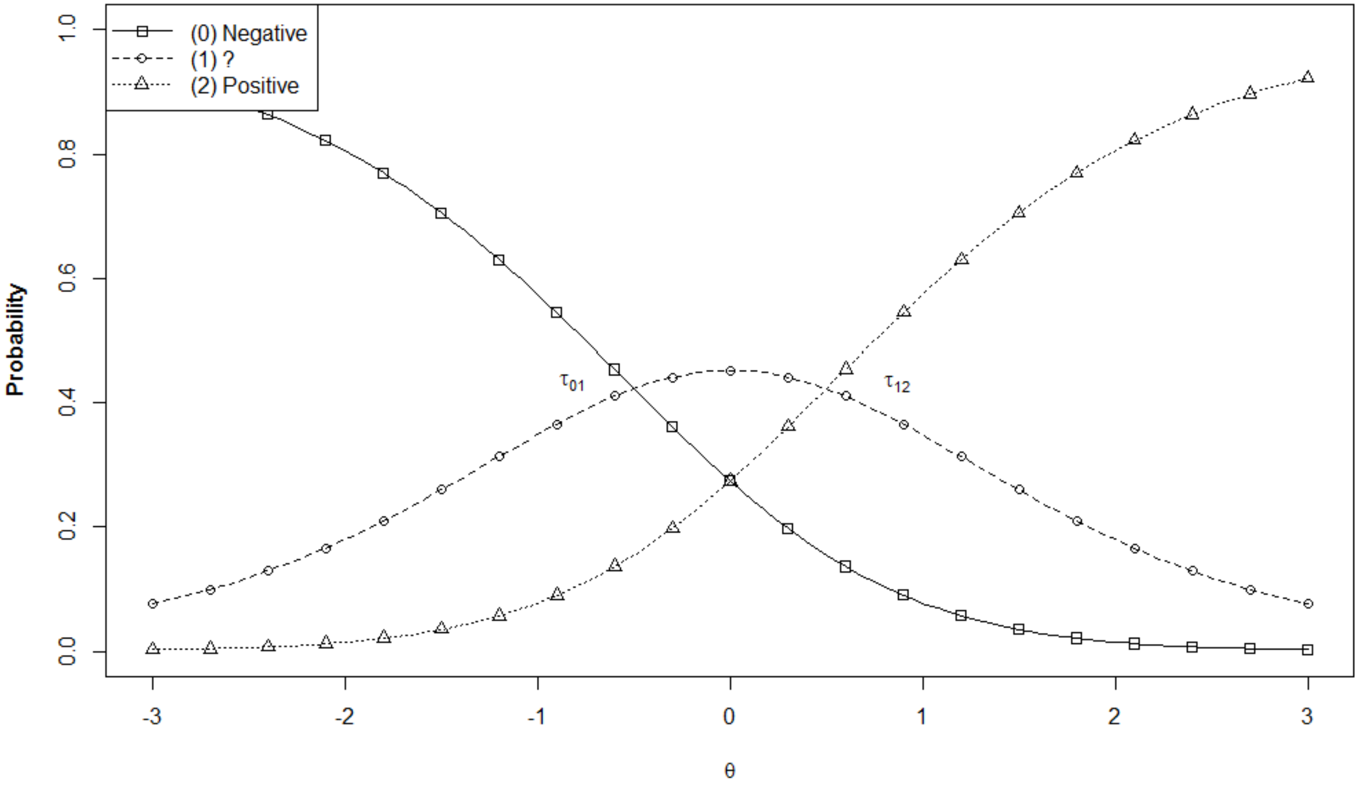
- What we assume…
  - That it represents the mid-point of the latent trait standing.
- How they are used…
  - Highly variable
  - Could represent the mid-point, uncertainly, would prefer not to say, don’t understand the question…
- A number of studies have looked at the issue, but none have done so:
  - In a large representative sample
  - Using an entire omnibus measure of personality
  - With optimal methods
  - And considered the impact of practical scoring algorithms

What we did…

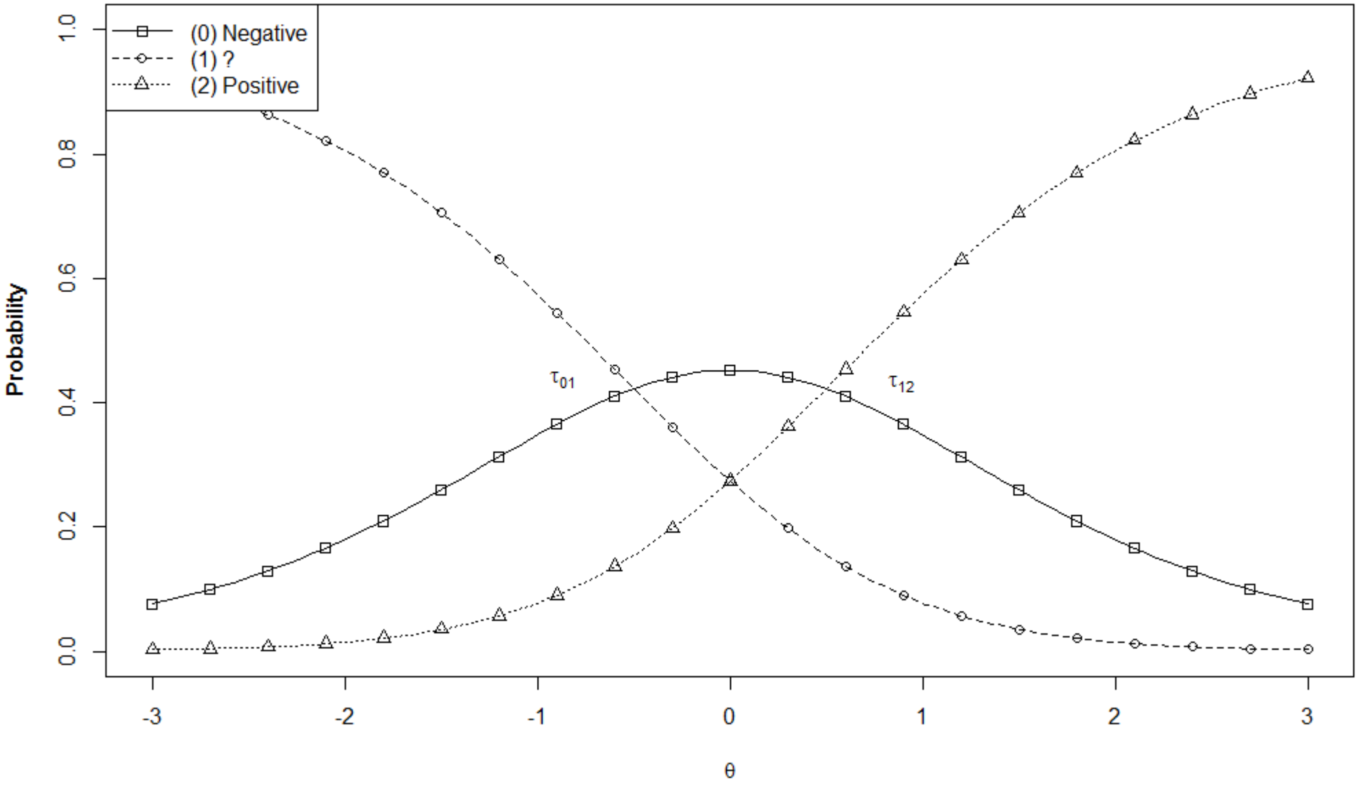
- **Data:**
  - US standardization sample of the 16 Personality Factor Questionnaire, version 5 (16PF5; n=10,261)
  - UK standardization sample of the 16PF5 (n=1,212)

- **Analysis:**
  - Approximate uni-dimensionality based on single factor CFA's
  - Compared fit of Nominal Response Model (NRM; Bock, 1972) to the Graded Partial Credit Model (GPCM; Muraki, 1992)
  - Investigated category and threshold ordering based on NRM

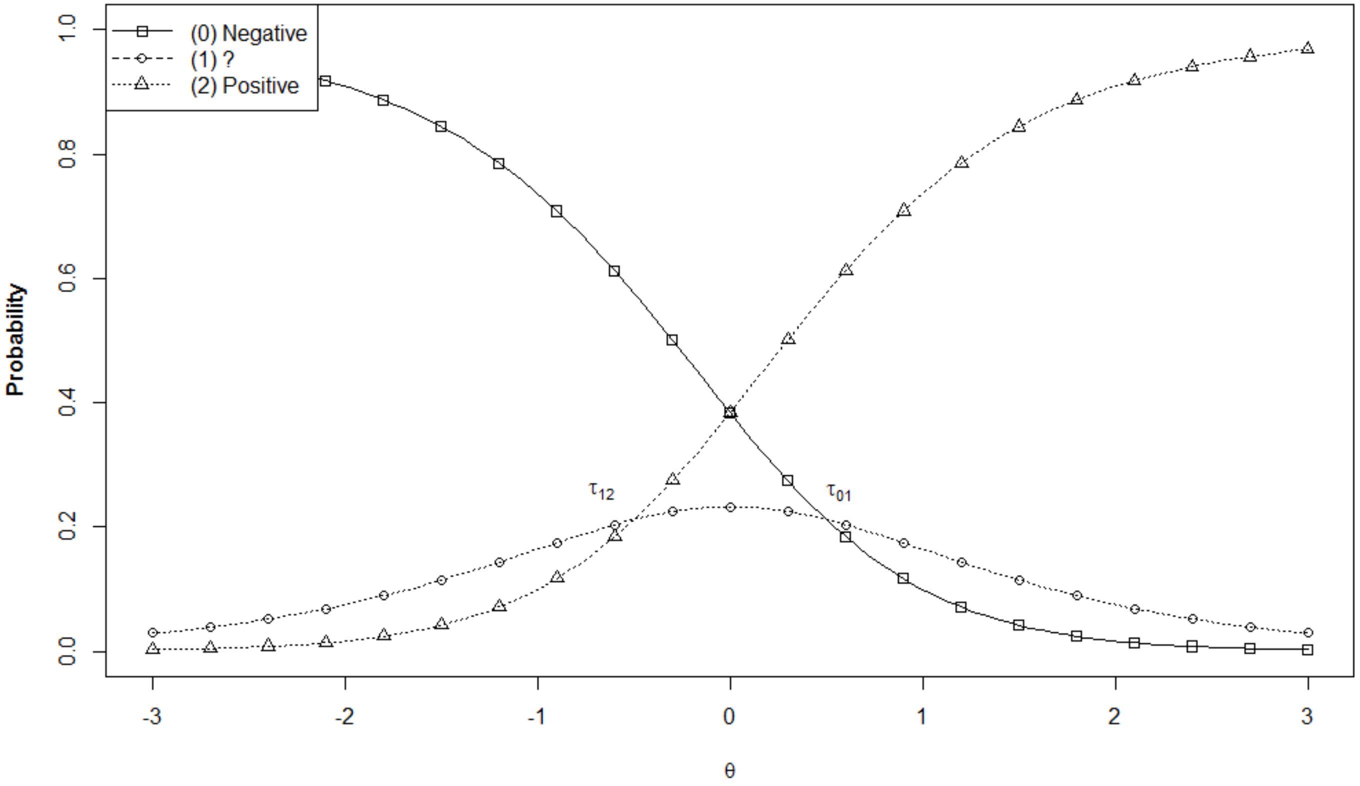
Ideal item response curves



Category disordering



Threshold disordering



US Results

Table 2. Nominal response model (NRM) model results for U.S. standardization sample.

Scale	Label	No. items	Mean ? endorsement rate	Results from NRM: Items where ? behaves as lowest response option	Items where ? behaves as highest response option	Items with thresholds out of order	NRM versus GPCM fit comparisons: $\Delta\chi^2$	df	ΔAIC	ΔBIC	$\Delta saBIC$
A	Warmth	11	.08	8	—	3	560.27	11	−538.27	−458.67	−493.63
C	Emotional Stability	10	.07	—	—	10	270.81	10	−250.81	−178.45	−210.23
E	Dominance	10	.09	—	—	10	138.90	10	−118.90	−46.54	−78.31
F	Liveliness	10	.10	—	—	10	59.78	10	−39.78	32.58	0.80
G	Rule Consciousness	11	.09	—	—	11	455.23	11	−433.23	−353.63	−388.59
H	Social Boldness	10	.09	—	—	10	55.15	10	−35.15	37.21	5.43
I	Sensitivity	11	.10	—	—	11	40.75	11	−18.75	60.85	25.89
L	Vigilance	10	.14	—	—	10	30.72	10	−10.72	61.65	29.87
M	Abstractness	11	.12	—	—	11	146.77	11	−124.77	−45.17	−80.13
N	Privateness	10	.09	—	—	10	38.16	10	−18.16	54.20	22.42
O	Apprehension	10	.08	—	—	10	118.61	10	−98.61	−26.25	−58.03
Q1	Openness to Change	14	.10	14	—	—	619.47	14	−591.47	−490.17	−534.66
Q2	Self-Reliance	10	.11	—	—	10	77.36	10	−57.34	15.01	−16.77
Q3	Perfectionism	10	.07	—	—	10	66.55	10	−46.55	25.82	−5.96
Q4	Tension	10	.09	—	—	10	40.73	10	−20.73	51.64	19.86

Note. GPCM = generalized partial credit model; AIC = Akaike's information criterion; BIC = Bayesian information criterion. Negative values of ΔAIC, ΔBIC, and $\Delta saBIC$ indicate that the NRM fits better than the GPCM.

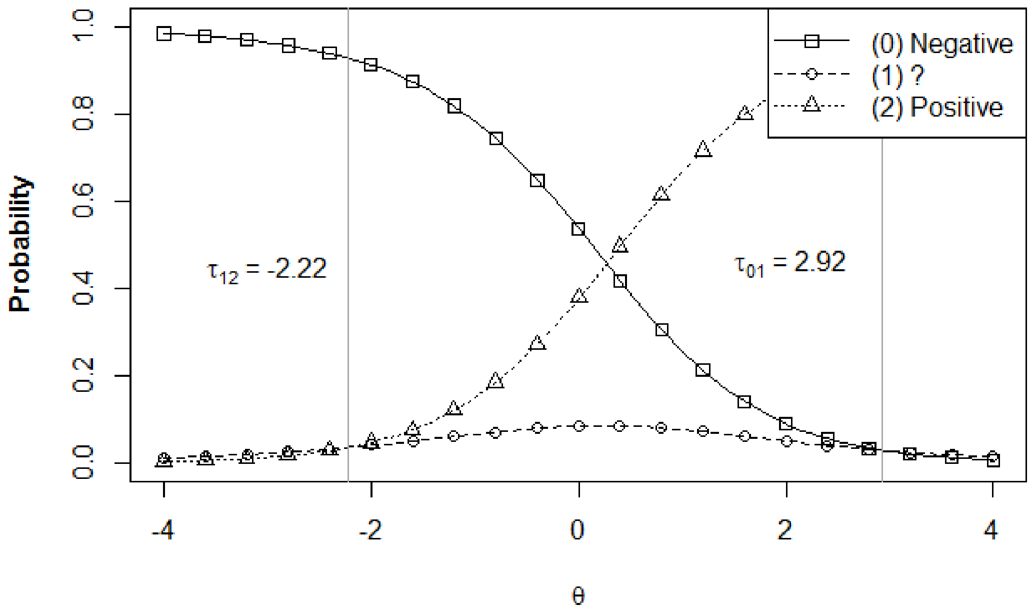
UK Results

Table 3. Item response theory model results for UK standardization sample.

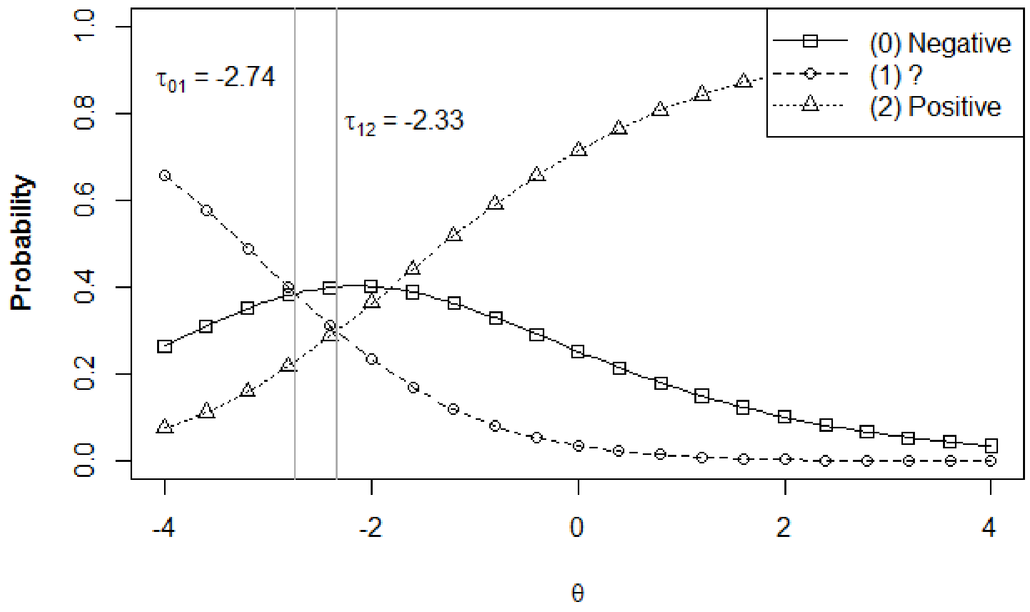
Scale	Label	No. items	Mean ? endorsement rate	Results from NRM: Items where ? behaves as lowest response option	Results from NRM: Items where ? behaves as highest response option	Results from NRM: Items with thresholds out of order	$\Delta\chi^2$	*df*	ΔAIC	ΔBIC	$\Delta saBIC$
								NRM versus GPCM model fit comparisons			
A	Warmth	11	.13	—	1	10	41.71	11	−19.71	36.40	1.456
C	Emotional Stability	10	.14	—	—	10	15.71	10	4.29	55.29	23.53
E	Dominance	10	.14	—	—	10	20.54	10	−0.54	50.46	18.70
F	Liveliness	10	.15	—	—	10	14.52	10	5.48	56.48	24.72
G	Rule Consciousness	11	.12	11	—	—	455.23	11	−433.23	−353.63	−388.59
H	Social Boldness	10	.13	—	—	10	19.13	10	0.87	51.87	20.10
I	Sensitivity	11	.13	—	—	11	40.75	11	−18.75	60.85	25.89
L	Vigilance	10	.18	—	10	—	261.59	10	−241.56	−190.56	−222.34
M	Abstractness	11	.19	—	—	11	146.77	11	−124.77	−45.17	−80.13
N	Privateness	10	.13	—	1	9	59.36	10	−39.36	11.64	−20.12
O	Apprehension	10	.13	—	10	—	276.24	10	−256.24	−205.24	−237.01
Q1	Openness to Change	14	.14	14	—	—	645.22	14	−617.21	−545.81	−590.28
Q2	Self-Reliance	10	.13	—	—	10	17.53	10	2.47	53.47	21.71
Q3	Perfectionism	10	.12	—	—	10	38.79	10	−18.79	32.22	0.45
Q4	Tension	10	.11	—	3	7	60.09	10	−40.09	10.91	−20.86

Note. NRM = nominal response model; GPCM = generalized partial credit model; AIC = Akaike's information criterion; BIC = Bayesian information criterion. Negative values of ΔAIC, ΔBIC, and $\Delta saBIC$ indicate that the NRM fits better than the GPCM.

Example Items

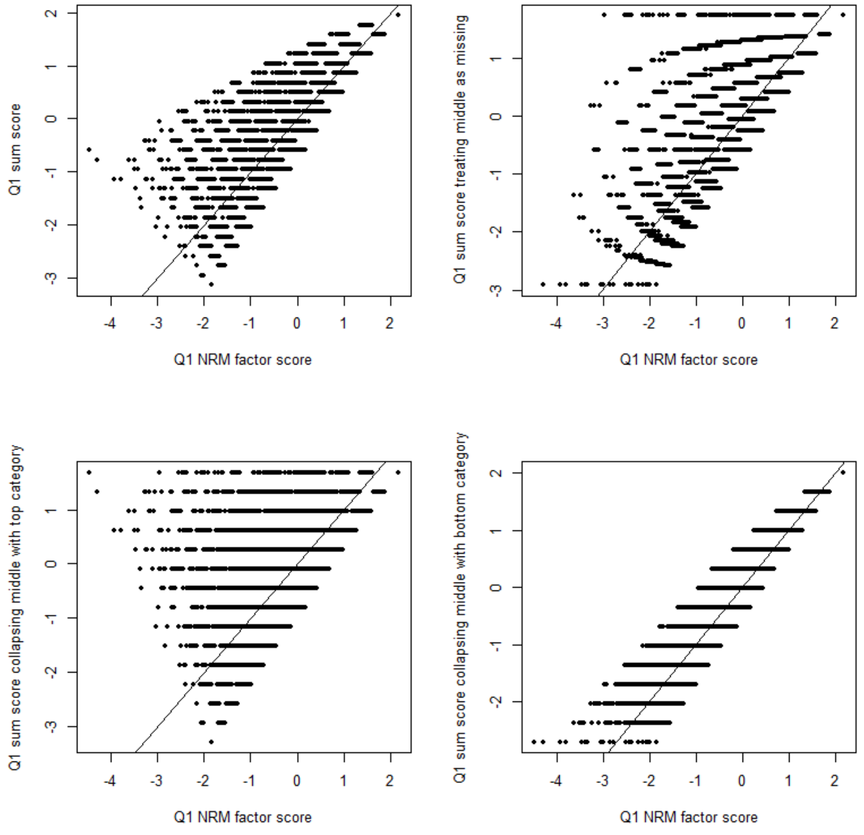


Item 1 of Sensitivity (I) in US sample



Item 1 Openness to Change (Q1), US sample

Practical Impact



Scores standardized for comparison:

NRM score <-> Sum; *r* = .83

NRM score <-> Missing data; *r* = .77

NRM score <-> Collapse with top; *r* = .62

NRM score <-> Collapse with bottom; *r* = .95

Practical Impact

Table 5. Percentages of participants selected by nominal response model factor scores who were also selected by simpler scoring methods.

Selection criterion	Sum score	Treat middle category as missing	Collapse middle category with lowest category	Collapse middle category with highest category
Top 100	2%	16%	29%	0%
Top 500	25%	35%	55%	14%
Top 1,000	43%	52%	69%	30%
Bottom 100	100%	45%	100%	42%
Bottom 500	69%	43%	80%	38%
Bottom 1,000	71%	65%	82%	48%

Epidemiology

(or something quite similar!)

Introducing the Lothian Birth Cohort 1936

- LBC1936 is a longitudinal study of ageing.
  - Surviving members of Scottish Mental Survey 1947.

- Wave 1 (n=1091), mean age 69.5 (SD 0.8)
  - Medical history, psychometric measures, cognitive testing, physical exam, blood samples.

- Wave 2 (n=866), mean age 72.5 (SD 0.7)
  - Repeat testing from wave 1. Additionally underwent brain scans.

- For full sample details see (Deary et al. 2007; Deary et al. 2011; Wardlaw et al. 2011).

Allostatic load and stress response

- When completing my post-doc in CCACE published on AL.

- **Allostasis:** Maintaining stability through change.
  - Homeostasis mentioned by Ram et al.

- **Allostatic Load:** Physiological cost of exposure to a stressor.

(Ganzel et al. 2010; McEwen & Wingfield, 2003).

AL in the literature

- Del Giudice et al. (2012) – “Adaptive Calibration Model”
- Chen & Miller (2012) – “Shift-and-Persist”

- Common features:
  - Key developmental periods in defining biological response to stress.
    - Early life stress (low SES, Adversity etc.) may be a major source of allostatic load.
  - Key roles of the sympathetic, parasympathetic and HPA-axis systems.
  - Individual differences may act as a buffering mechanism.
  - Effects of allostatic overload seen in numerous biomarkers including brain tissue integrity.

AL in the literature

- SES/Environments -> Health & Biomarkers of Allostatic Load
  - E.g. Dowd et al. (2009); Szanton et al. (2005); Seeman et al. (2010)
- Biomarkers of AL -> Health:
  - E.g. Carlson & Chamberlain (2005); Juster et al. (2010)
- Importance of key developmental periods:
  - E.g. Wilkinson & Goodyear (2011); Howell & Sanchez (2011)
- Role of personality and individual differences:
  - E.g. Chapman et al. (2009); Nabi et al. (2008)

Measuring AL

- Across studies (e.g. Szanton et al. 2005; Chen et al. 2012) suggest 8 areas and markers:

    1. Sympathetic Nervous System: Nor/Epinephrine
    2. Parasympathetic Nervous System: High/low frequency heart rate
    3. HPA-Axis: Cortisol, DHEA.
    4. Inflammation: C-reactive Protein, Iterleukin-6, fibrinogen.
    5. Cardiovascular: Systolic and diastolic BP, heart rate.
    6. Glucose Metabolism: Glucose, Insulin resistance.
    7. Lipid Metabolism: High/low-density lipoprotein, BMI, waist-hip ratio, triglycerides.
    8. Organ function: Creatinine, Respiratory peak flow.

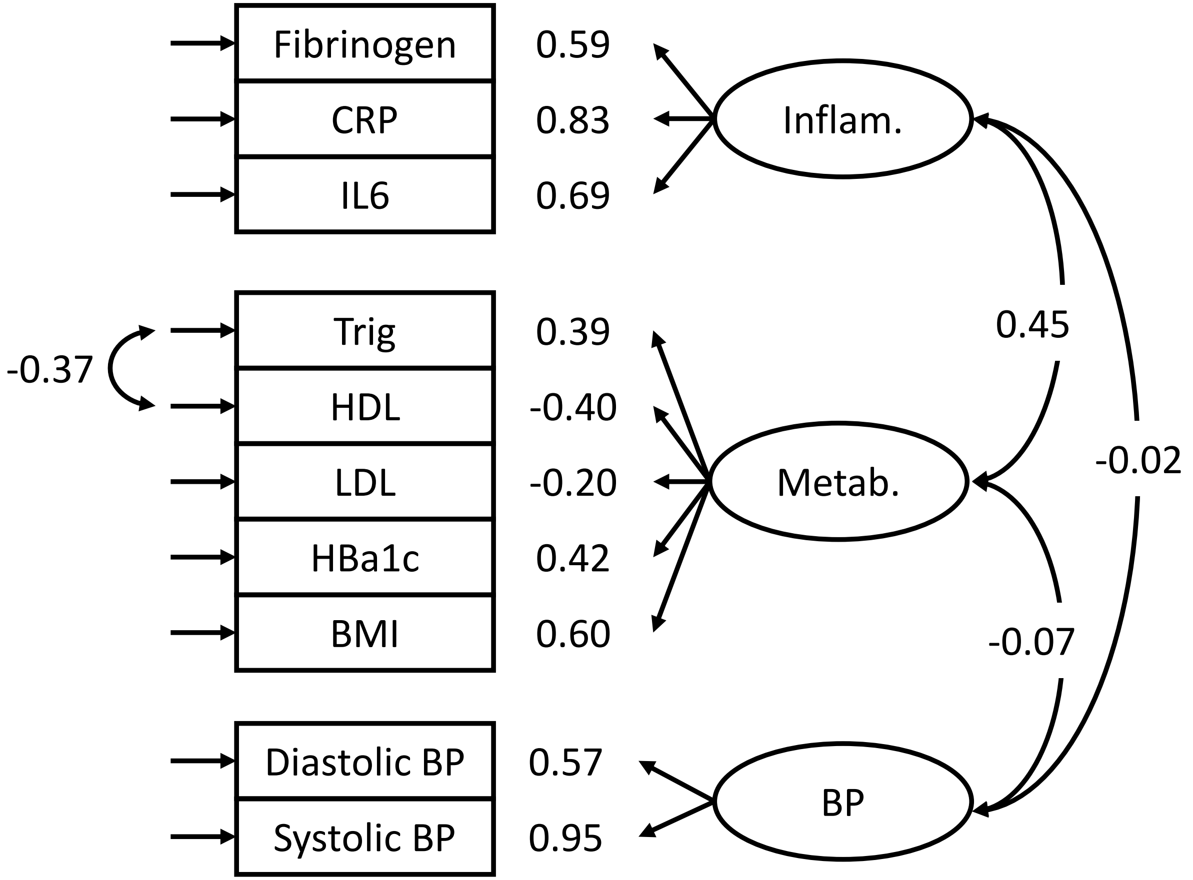
- Seeman et al. (2010) produced a second order CFA model using 5 areas as factors and 17 biomarkers.

Replicating the measurement model

- Based on wave 2 data from the LBC1936

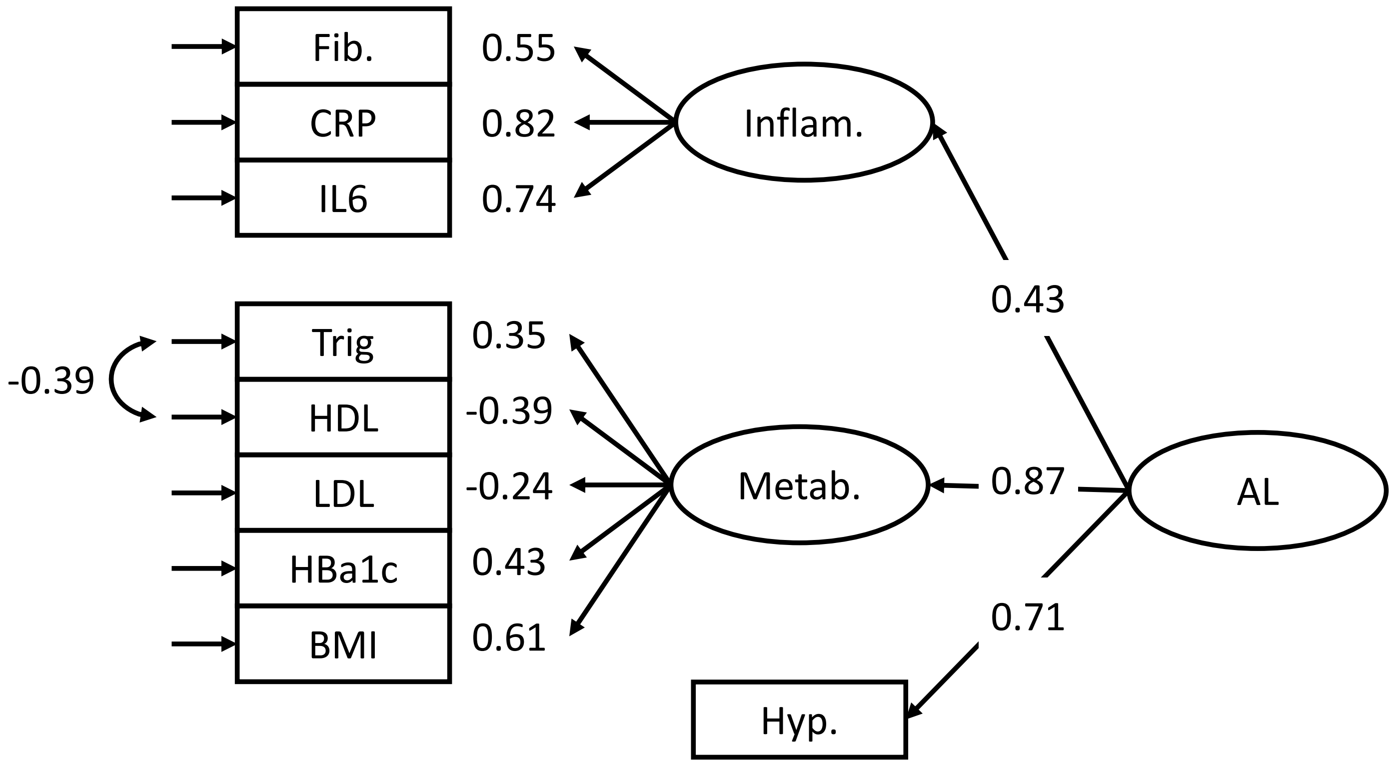
- Allostatic load measures:
  - **Metabolic Markers:** Triglycerides (Trig), High density Lipoprotein (HDL), low density lipoprotein (LDP), HBa1c & Body-Mass Index (BMI).
  - **Inflammation Markers:** Fibrinogen, C-reactive Protein (CRP) & Interlukin-6 (IL6)
  - **Blood Pressure:** Mean sitting systolic and diastolic blood pressure (SBP & DBP), & ortostatic hypotension (orto).

Replicating the measurement model



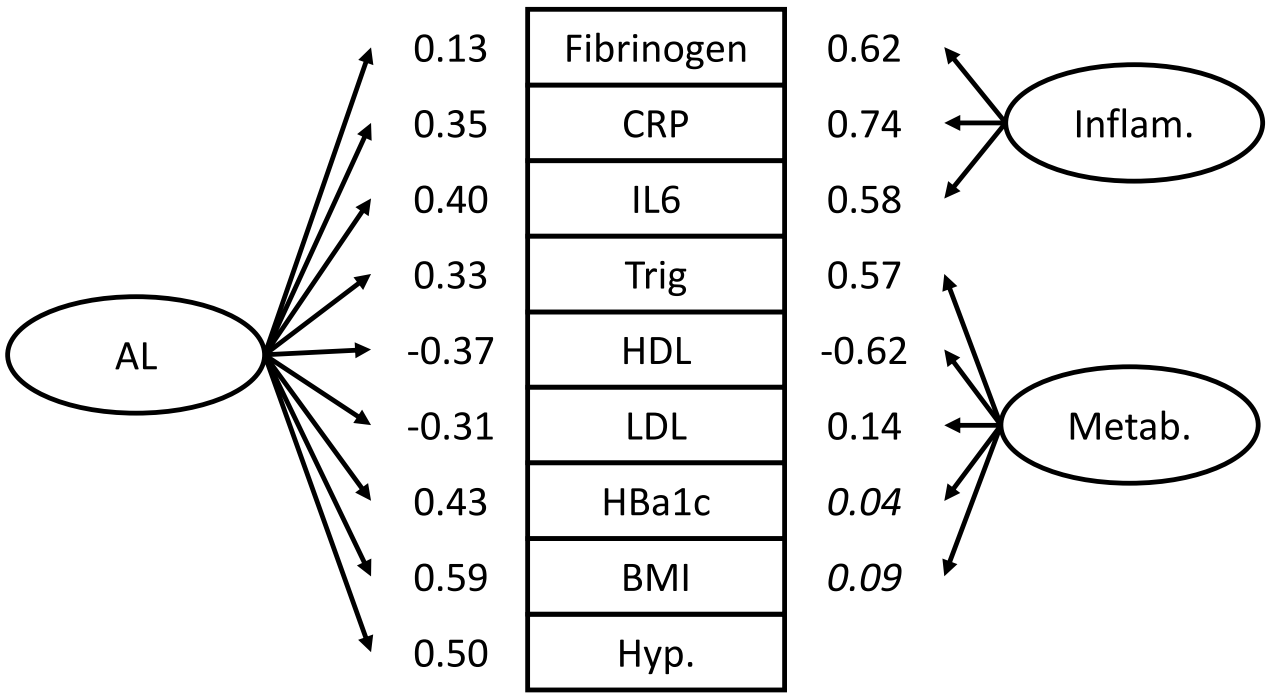
Booth, Starr & Deary (2013). Modeling multisystem biological risk in later life: Allostatic load in the Lothian birth cohort study 1936. *American Journal of Human Biology, 25,* 538-543

Replicating the measurement model



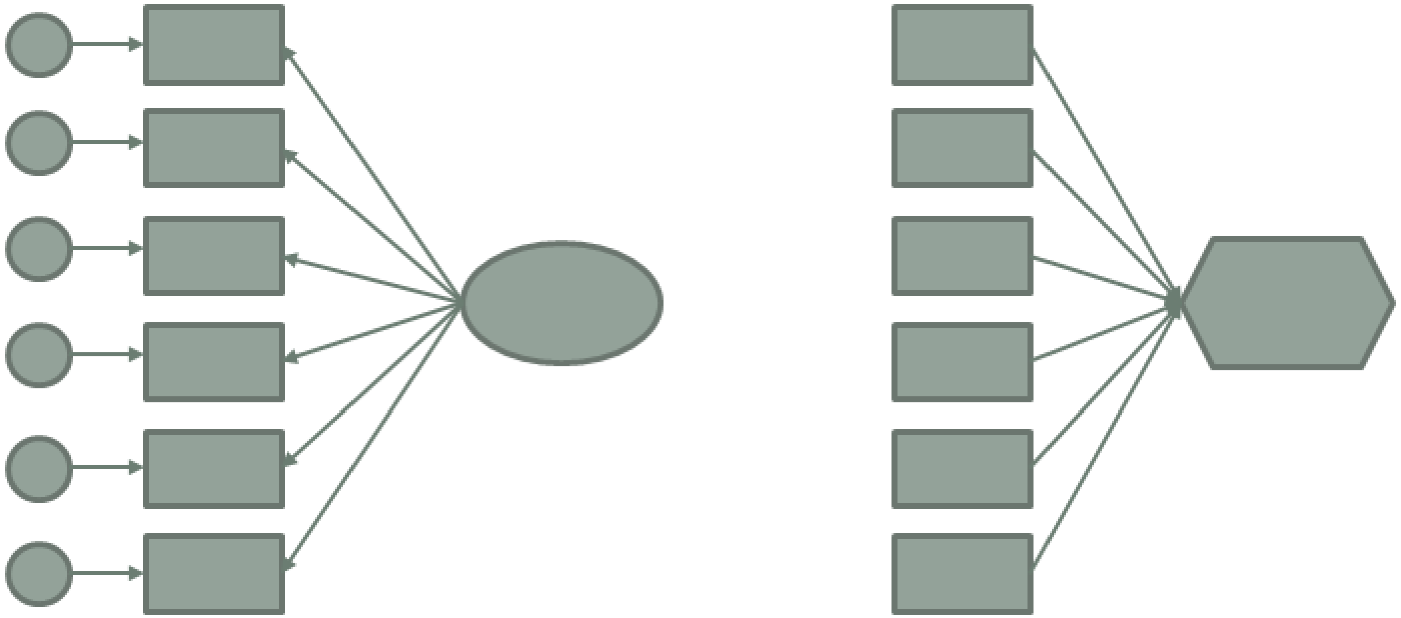
Booth, Starr & Deary (2013). Modeling multisystem biological risk in later life: Allostatic load in the Lothian birth cohort study 1936. *American Journal of Human Biology, 25,* 538-543

Replicating the measurement model



Booth, Starr & Deary (2013). Modeling multisystem biological risk in later life: Allostatic load in the Lothian birth cohort study 1936. *American Journal of Human Biology, 25,* 538-543

Measuring AL: Factor or composite?



Common methods of assessment

- Sample specific cut-points (deciles, quartiles)
- Clinical cut-points
- Composite scores (summed z-scores, PCA)
- Latent models (higher-order, bi-factor)

Measurement of AL

	AL-bifactor	AL-PCA	AL-z	AL-dec	AL-quart	AL-clin
AL-bifactor	1.00					
AL-PCA	0.58	1.00				
AL-z	0.47	0.72	1.00			
AL-dec	0.39	0.65	0.59	1.00		
AL-quart	0.45	0.69	0.67	0.68	1.00	
AL-clin	0.42	0.65	0.80	0.60	0.78	1.00

Booth (Unpublished). Supplementary allostatic load analyses in the LBC1936.

AL and early life adversity

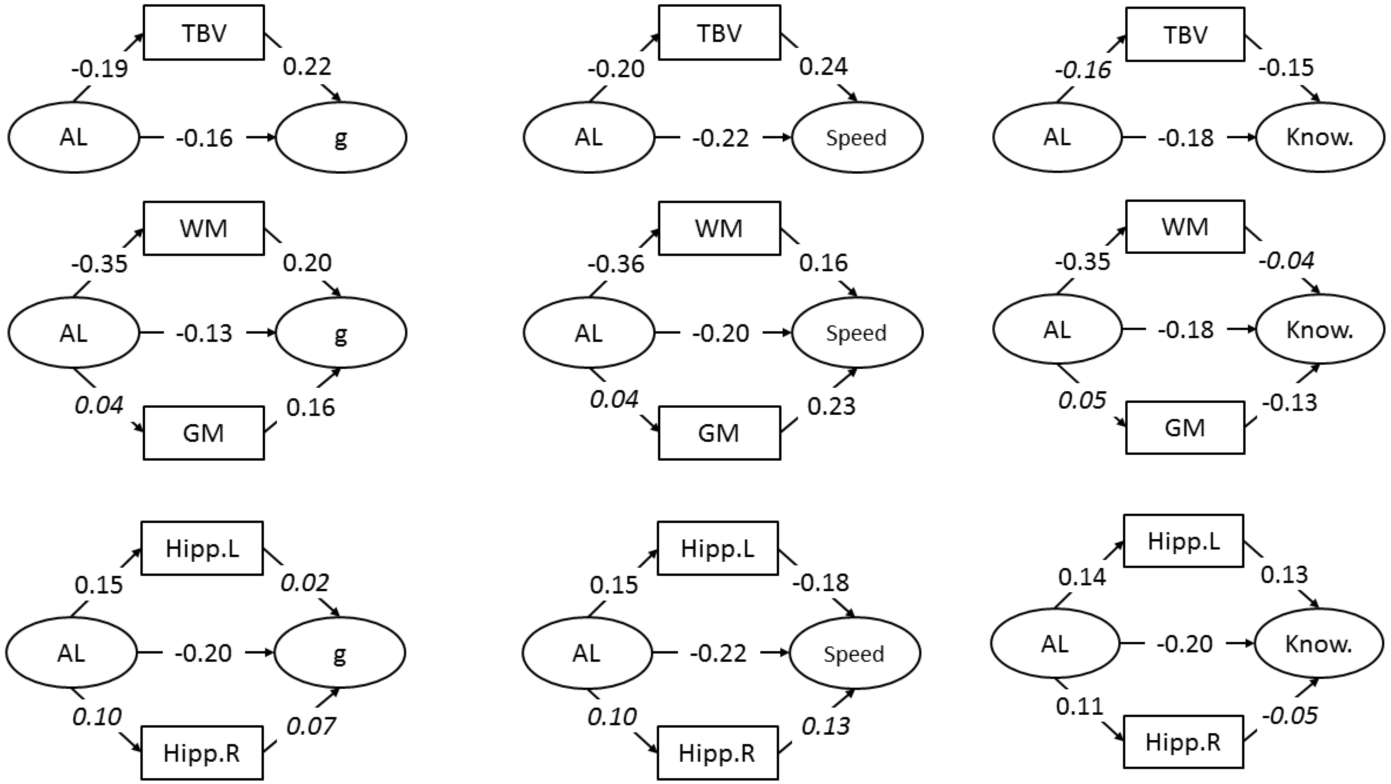
	AL-bifactor	AL-PCA	AL-z	AL-dec	AL-quart	AL-clin
Adversity	0.09	0.09	0.10	0.12	0.11	0.04
	0.11	0.08	0.09	0.10	0.11	0.02

- Childhood Adversity (Johnson et al. 2010) scores based on living environment at age 11 years:
    - People/rooms in house ratio.
    - People per toilet.
    - Indoor toilet (yes/no).
    - Fathers occupational class.

Booth (Unpublished). Supplementary allostatic load analyses in the LBC1936.

Other AL Studies

- Medication of AL-> cognitive ability association by brain volumes
- AL white matter tract integrity associations using TBSS: assessing localization and lateralization of effects.
- Moderated mediation of depression and AL on cognitive decline by APOE status
- AL and methylation age
- Optimal measurement of AL by criterion association



Booth et al. (2015). Association of allostatic load with brain structure and cognitive ability in later life. *Neurobiology of Ageing, 36,* 1390-1399

A methodological interlude

The importance of time when integrating social and biological

Multiple timeframe studies of biosocial processes

- The broad argument is seated in what might now be considered classical distinctions for longitudinal studies (Nesselroade, 1991):

  - **Change:** Longer term processes usually studied within traditional panel or cohort studies.

  - **Variability:** Shorter term processes or fluctuations, now becoming studied with experience sampling methodologies.

- But what happens when we extend to multiple timeframes, and pose hypotheses about the relations between effects at these different time frames?

AL and multiple time frames

Process	Time scale
Allostatic load (accumulation)	Life long
Environmental influence	Life long but periodic – we move, situations change
Traits (personality, IQ, anxiety)	Argued long term stability, but evidence of high within person variability
Biomarkers	Momentary response cycles (cortisol)

Optimal time lags

"With impoverished theory about issues such as when events occur, when they change, or how quickly they change, the empirical researcher is in a quandary. Decisions about when to measure and how frequently to measure critical variables are left to intuition, chance, convenience, or tradition. None of these are particularly reliable guides" (Mitchell & Jones, 2001; cited in Dormann & Griffin, 2015, p.1)

Optimal time lags

"Collins (2006) argues that an effective longitudinal design depends on capturing the theoretical process that is consistent with the temporal change being investigated." (Dormann & Griffin, 2015)

- Cole and Maxwell (2009):
  - In discussing clinical reseach note that when the time lag selected is inappropriate, common analytic methods for panel designs lead to under-estimation of true effects.
- Cohen, Cohen, West & Aitken (2003) - classic regression book:
  - As the lag becomes largely, the estimate of the effect decreases.

Individual differences and health

Where I have ended up!

Personality and Physical Health Review

- Lots of studies of simple associations.
    - Some from very large samples; some using meta-analysis etc.

- Few studies looking at change.
    - Only one we found looked at the relationships between personality, behaviour change and health change (Takahasi et al. 2013, J. Personality)

- Mediation studies rarely considered the strength of mediation effect.

Murray & Booth (2015). Personality and Physical Health. *Current Opinion in Psychology, 5,* 50-55

Personality and Physical Health Review

- Moderation results are inconsistent
  - E.g. evidence for healthy neuroticism.
  - Some evidence for environmental moderators

- Item, facet or factor level effects?

- Interventions and intensive longitudinal studies

Personality, health and brain integrity

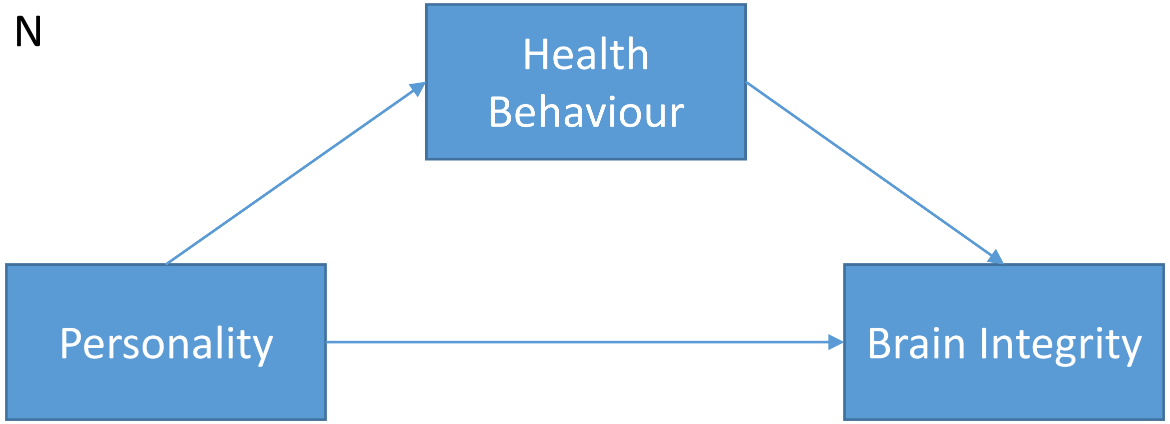
- Here we operationalised global brain integrity in old age as a measure of health.
  - Moved away from functional explanations for personality <-> brain associations.

- Global measures here refers to things like total brain volume, white matter volume, average tract integrity.
  - Some exceptions (e.g. clinical).
  - Not talking here about fMRI

Booth, Mottus et al. (2015). Personality, health, and brain integrity: The Lothian Birth Cohort Study 1936. *Health Psychology, 33,* 1477-1486

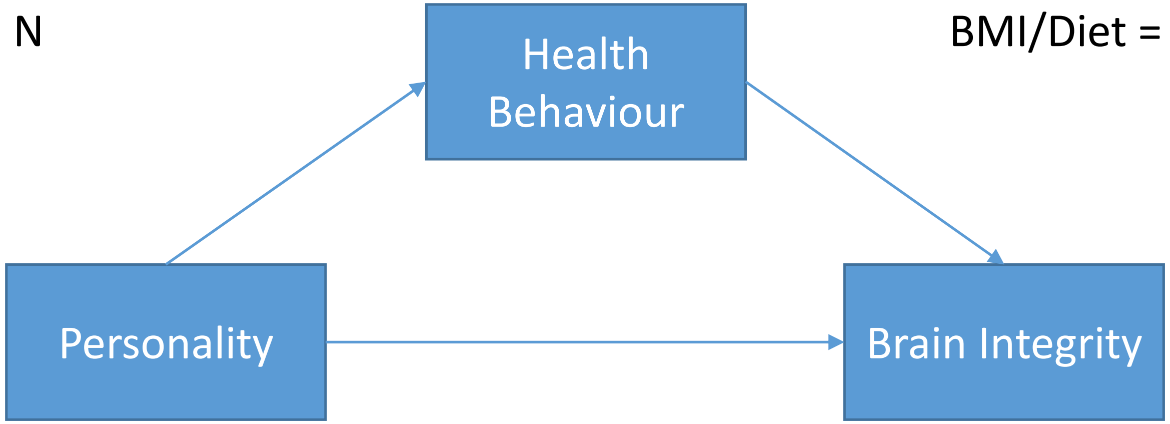
Proposed mechanism

Smoking = C
Diet = C, A, E, O, N
BMI/Weight = C
Physical Activity = C, E, N
Alcohol Intake = C



Proposed mechanism

Smoking = C
Diet = C, A, E, O, N
BMI/Weight = C
Physical Activity = C, E, N
Alcohol Intake = C



Physical Activity = WML, Atrophy
Smoking = WM/tissue integrity
Alcohol = WM/tissue integrity
BMI/Diet = Regional volume, WM integrity

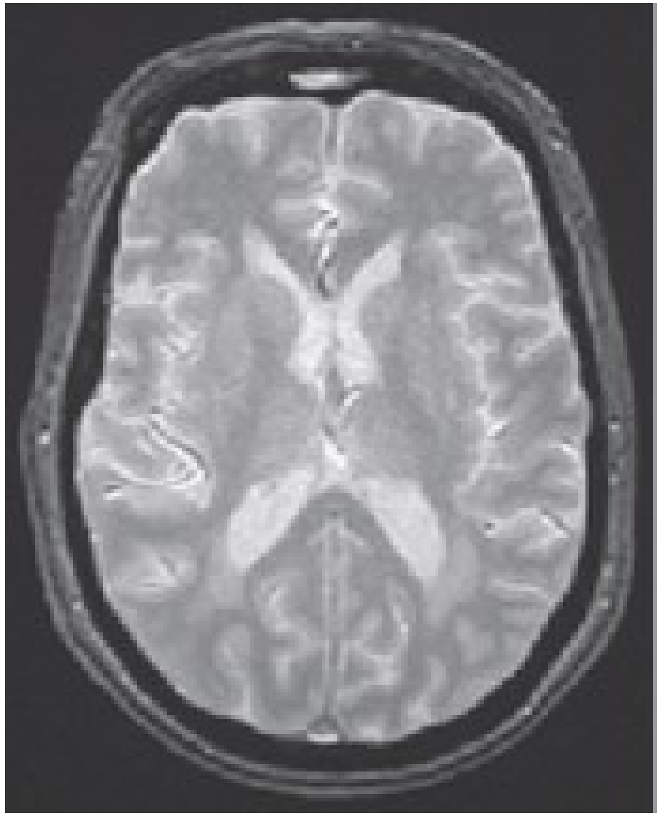
Sample and data

- LBC1936: n’s between 529 to 565

- Personality (Wave 1): IPIP five factor model inventory
  - 50 items, 5-point likert scale, 10 items per factor.
  - Standardized item average used here.

- Health behaviours:
  - Pack years ((Average No. of cigarettes per day*years as a smoker)/20); alcohol consumption (per week); physical activity (single self-report item); health aware diet; BMI.
  - Composite of all above
    - PCA: loadings between 0.29 to 0.42, 13% variance explained

Atrophy



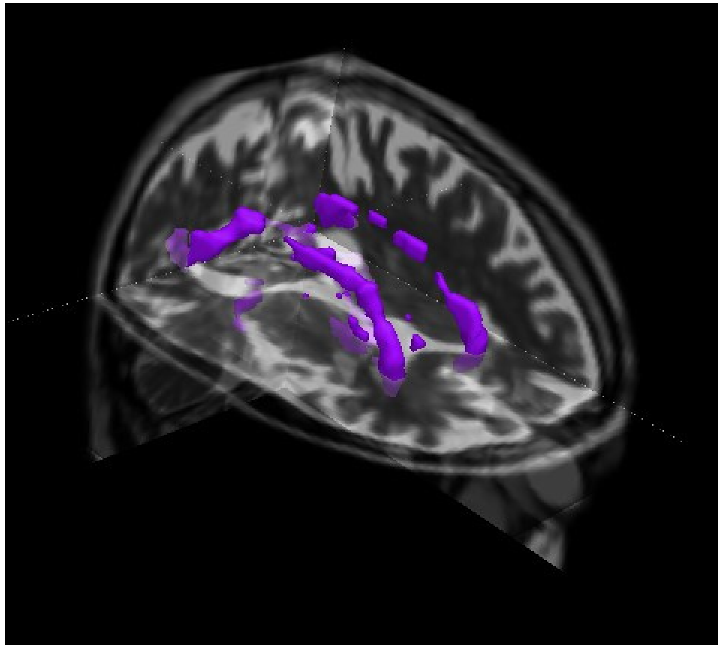
- Under the assumption that intracranial volume (ICV) peaks at maximal brain growth and remains stable, we computed the ratio of total brain volume (TBV) to ICV as;

- Atrophy = (1-(TBV/ICV))*100

- Figure: Example slice from T2*Weighted image showing the outline of brain tissue and ICV.

(Valdés Hernández et al., 2010; Valdés Hernández et al, 2012)

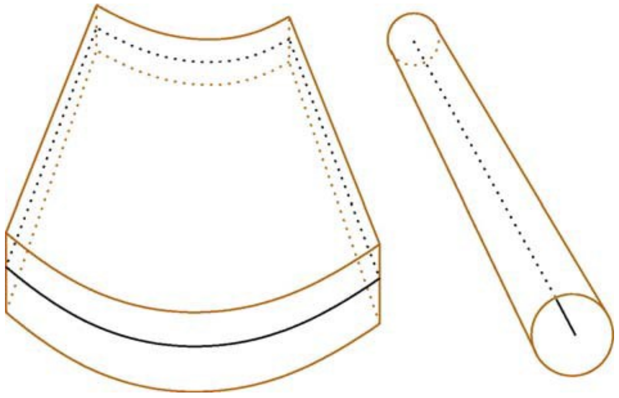
White Matter Lesion (WML) Volume

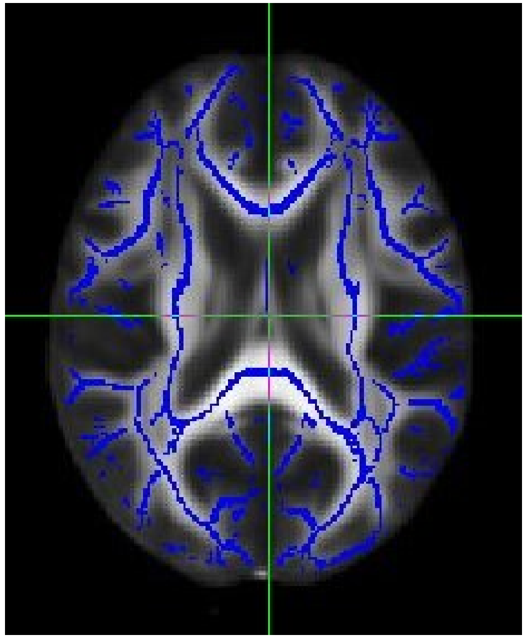


(Valdés Hernández et al., 2010)

- WML volume was estimated using the MCMxxxVI method by fusing T2*weighted and FLAIR images.
- Manual visual quality check for false positives and old infarcts.
- Figure: WML (purple) in 3D superimposed into a T2-weighted MRI volume (type of structural MRI).

Average White Matter Tract (WMT) Integrity





- **Fractional Anisotropy (FA):**
- Metric of the degree of directional dependence on the movement of water molecules in white matter tracts.
- Figure: Tract structures with demarcation of the tract centre. Target of TBSS projection.

- **TBSS:**
- Average FA across voxels of study specific white matter skeleton.
- Figure: TBSS mean white matter skeleton (Axial Slice, Blue).

Analysis Strategy

- Correlation and regression to assess uni- and multivariate associations between personality, mediators and brain integrity measures.
- Mediation analysis using “mediation” package in R (Imai, Keele & Tingley, 2010)
  - Counterfactual causal mediation framework.
  - Estimates direct, indirect and mediated effects of a “treatment” (personality) via a mediator (health behaviours) to an outcome (brain integrity) in the presence of “pre-treatment confounders” (age and sex).
- Stability checks.

Results: Personality <-> brain integrity

	N	E	I	A	C
WMT FA	-0.11^{*}	0.07	0.06	0.14^{**}	0.20^{***}
Atrophy	-0.04	0.04	-0.03	-0.18^{***}	-0.13^{**}
WML Volume	0.03	0.01	-0.04	0.03	-0.11^{**}

- Multiple regression models controlling for age and sex, only Conscientiousness was significantly associated:
  - WMT FA = .16, $p<.001$
  - Atrophy = -.11, $p<.01$
  - WML = -.10, $p<.05$

Results: Personality <-> Health behaviours

	N	E	I	A	C
Pack-Years	0.13**	-0.06	-0.05	-0.13**	-0.10*
Physical Activity	-0.12**	0.04	0.14**	-0.03	0.13**
Diet	-0.02	-0.05	0.12**	0.20***	0.08
Alcohol Units	-0.05	0.11**	-0.02	-0.20***	-0.07
BMI	0.03	0.04	-0.08	0.00	0.10*
Health Composite	0.10*	0.01	-0.13**	-0.15***	-0.14**

Proportion of direct effect mediated by C

	WMT FA Est.(CI)	Atrophy Est.(CI)	WML Volume Est.(CI)
Physical Activity	**.061 (.005 to .168)**	**.137 (.028 to .622)**	**.119 (.014 to .475)**
Pack-years	.064 (-.025 to .181)	.063 (-.024 to .222)	.055 (-.025 to .271)
Health-Aware Diet	.019 (-.010 to .095)	.094 (-.034 to .413)	-.032 (-.150 to .037)
Alcohol Intake	.009 (-.016 to .049)	.114 (-.013 to .458)	.002 (-.076 to .084)
BMI	.056 (-.004 to .182)	-.007 (-.135 to .078)	.033 (-.055 to .183)
Health Composite	**.107 (.022 to .259)**	**.177 (.049 to .798)**	.072 (-.022 to .365)

Stability Checks

- Re-ran the models using a single SEM.
  - Questions of composite factor.
- Considered the associations of single C items with brain integrity measures.
  - Is it driven by single items?
- Conducted a specificity check (Mõttus, et al. Submitted).
  - Are the associations larger than random scales?
- Re-ran correlations with brain integrity using NEO-FFI wave 1.
  - Are the findings measure specific?

Discussion & Limitations

- Consistent C association supporting “health” claim.
  - Explicit test of our proposed mechanism for the association.
- Larger mediation effects of the health composite point (tentatively) at a cumulative process.
  - But all effects were small.
  - General concern over quality of health behaviour measures.
- Stability checks point to robustness of the solutions.
- Large sample of healthy ageing individuals.

Interesting reviewer comments

"…basic association results seem to be a valuable scientific contribution. However, I do not find the framing of this manuscript compelling….In my opinion, the study can have some impact if stripped of the mediation part."

"…the bigger contribution of this paper for me is not the data as such but the placement of this particular data in the broader debate of what it means to be a neural correlate of personality. Specifically the authors oppose the view that correlations between (global) brain measures and personality do not, by that operation, lend themselves for to a functional interpretation informing on the neurological basis of traits. This. for me, raises this paper, with very expected results, a degree."

Looking forward

Multiple timeframes and experience sampling

Wrapping up

- Broad area is hugely interesting, very rewarding, methodologically challenging.
  - For me that is the recipe for interest
- Add to this the potential for both individual and societal impact through health improvement.
- I am think there is huge potential of multiple time frame studies in this area.
  - As well as intensive longitudinal measures, like experience sampling, for individual level intervention.
- That is where is psychometrician is hoping to take his epidemiology.

Ambitious with my time slot…

- In the abstract I spoke of:
  - Bi-factor models: I have used these to partition variance in questions of the relation between specific cognitive abilities and brain imaging metrics.
  - Network models: I have one data collection complete where we applied multi-level vector autoregressive network analysis to study of work stress, personality and positive and negative affect before and after a time management intervention.
    - Results suggest little change in network structure
    - Follow on analysis to consider dual trajectories of stress, negative affect and the possibility of de-coupling after intervention.
- Also interested in recent discussions of parameterization of longitudinal models in derivative language….interesting idea…. “language of change”

Thank you very much.

Any questions?

(Reference PDF will be available)