Integrative Analysis of Longitudinal Studies on Aging

Integrative Analysis of Longitudinal Studies on Aging (IALSA):

Challenges and Requirements for Quantitative Harmonization

Scott M. Hofer & Andrea M. Piccinin

University of Victoria

The Integrative Analysis of Longitudinal Studies of Aging (IALSA) research network is supported by grants from the National Institutes of Health: AG026453 and Canadian Institutes of Health Research: 200910MPA Canada-UK Aging Initiative.

IALSA

Integrative Analysis of Longitudinal Studies on Aging

Collaborative Research Networks, Meta-Analysis, and Optimizing Future Studies

Andrea M. Piccinin and Scott M. Hofer

Remarkable national and international efforts have produced well over 40 major longitudinal studies of individuals age 50 and older with a significant cognitive assessment component. It is widely recognized that although longitudinal information is time and effort intensive to collect, it is required to address central questions in developmental research relating to intraindividual change and variation and to population inference condi-

and software for analysis of longitudinal data, provide unprecedented opportunities for describing and explaining aging-related changes and cross-process dynamics and for identifying influential factors associated with late life outcomes. Longitudinal studies permit the identification of change from a within-person baseline, thus enabling the identification of characteristics and antecedents that are potentially causally related and amenable to intervention.

Piccinin, A. M. & Hofer, S. M. (2008). Integrative analysis of longitudinal studies on aging: Collaborative research networks, meta-analysis, and optimizing future studies. In S. M. Hofer and D. F. Alwin (Eds.), *Handbook on Cognitive Aging: Interdisciplinary Perspectives (446-476).* Thousand Oaks: Sage Publications.

University of Victoria

NIA Workshop: Harmonization Strategies for Behavioral, Social Science, and Genetic Research

http://www.nia.nih.gov/sites/default/files/nia_bssg_harmonization_summary_version_2-5-20122.pdf

Appendix 4: Steps of Harmonization

Define the research question(s)

Prospective Harmonization	Retrospective Harmonization	
Achieve consensus on compatible study designs, measures and collection procedures	Catalogue study characteristics and database content	Identify common variables of interest
Ensure quality and consistency of common data collection	Evaluate harmonization potential of the participant studies	

Process study-specific data under a common data format and achieve quality control analysis

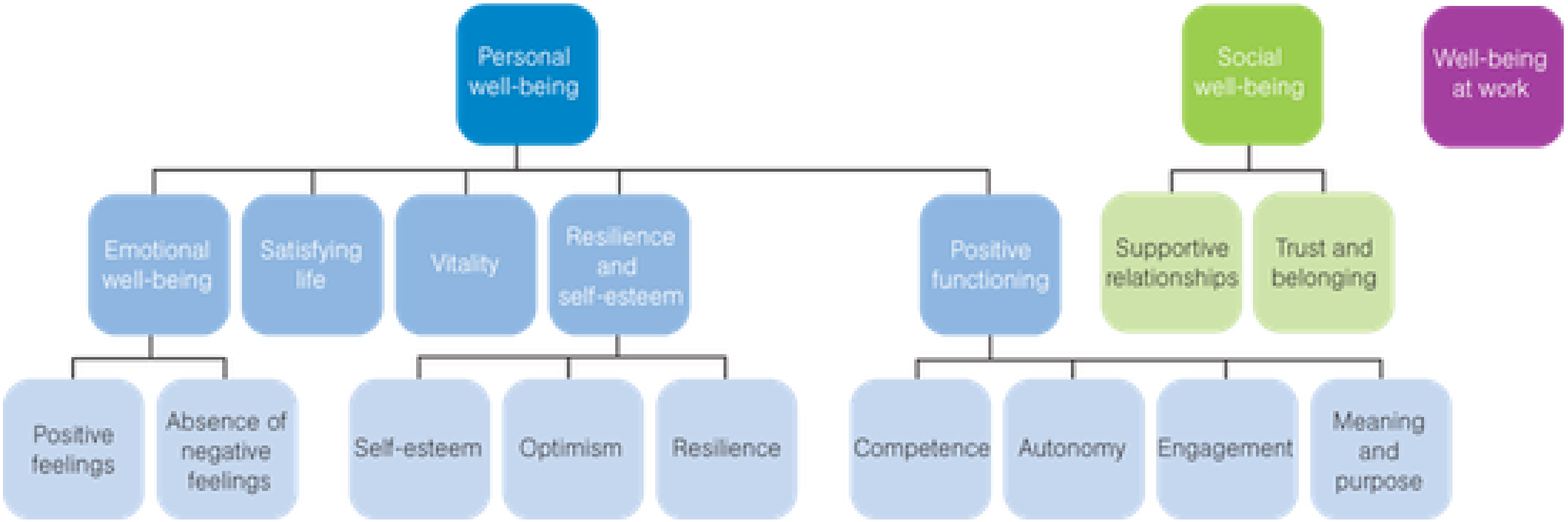
Integrate and, if relevant, transfer harmonized data

Achieve data analysis

Measurement Harmonization

- Aim: To understand complex interplay of genetic, lifestyle, environmental, and social factors associated with health and aging-related outcomes
  - Large samples are essential (Fortier et al., 2010)
  - Measurement: construct validity, reliability
  - Within-person designs
    - WP change →BP diffs; Birth cohort; Selective mortality; Age-heterogeneity confounding associations

National Accounts of Wellbeing



http://www.nationalaccountsofwellbeing.org

Measurement and Change

- Factorial invariance
  - Construct validity across age, ability, culture
- Reorganization of individual differences
  - Maturation/Aging; morbidity/mortality
- Baseline age heterogeneity
  - Confound w/mean trend (e.g., Hofer, Flaherty, & Hoffman, 2006); mortality selection

Integrative Data Analysis

From *Replication* to *Synthesis* to *Extension*

- **Sequential replication**: Evaluation of a published result on an independent set of data.
- **Meta-analysis (aggregate data)**: Combination of standardized effects from a set of published findings in order to estimate the general effect and to understand why studies differ in their results.
- **Coordinated replication**: Coordinated analysis (i.e., same statistical model) of individual data sets in ways that maximize comparability across study results.
- **Data Pooling**: Methods for combining individual-level data sets within a single analysis (i.e., individual patient meta-analysis, mega-analysis) permitting evaluation of both study-level and individual-level effects.
- **Generalized evidence synthesis**: Methods for combining data from multiple sources and analyzing models that cannot be evaluated in any single source of data (e.g., across age ranges, variable domains).

Hofer & Piccinin, 2009, Psychological Methods

Current limitations for meta- and pooled data analysis of longitudinal observational studies

- Challenges for Meta-Analysis
  - Paucity of published information on particular research questions
  - Complexity of differences across longitudinal studies
  - Different statistical analyses and limited reporting of results in published findings.
    - Summarizing regression slopes is challenging because of different measures (metrics) for X and Y and differences in statistical models (e.g., predictor sets) (Becker & Wu, 2007)
- Challenges for Pooled Data Analysis
  - Differences in measurements across studies (construct validity)
  - Insufficient basis for cross-study item-level calibration

Harmonization

- Goal: To make research results sufficiently comparable to obtain systematic answers to key questions, providing evidence for generalizability
- Levels of Harmonization
  - Measurements
  - Research Questions
    - (determinants, outcomes, mediators/moderators, confounders)
  - Statistical Models
- Harmonization permits synthesis of results
  - Account of how birth cohort, country/culture, attrition/mortality, historical period, etc. relate to differences across studies.

Piccinin, A.M., Muniz, G., Sparks, C., & Bontempo, D.E. (2011). An evaluation of analytical approaches for understanding change in cognition in the context of aging and health. *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences, 66B(S1)*, i35–i48, doi:10.1093/geronb/gbr038

An Evaluation of Analytical Approaches for Understanding Change in Cognition in the Context of Aging and Health

Andrea M. Piccinin,[1] Graciela Muniz,[2] Catharine Sparks,[1] and Daniel E. Bontempo[3]

[1]Department of Psychology, University of Victoria, Canada.
[2]MRC Biostatistics Unit, Cambridge, UK.
[3]Schiefelbusch Lifespan Institute, University of Kansas, Lawrence.

Objectives. In this article, we discuss the importance of studying the relationship between health and cognitive function, and some of the methods with which this relationship has been studied.

Methods. We consider the challenges involved, in particular operationalization of the health construct and causal inference in the context of observational data. We contrast the approaches taken, and review the questions addressed: whether health and cognition are associated, whether changes in health are associated with changes in cognition, and the degree of interdependency among their respective trajectories.

Results. A variety of approaches for understanding the association between cognition and health in aging individuals have been used. Much of the literature on cognitive change and health has relied on methods that are based at least in part on the reorganization of between-person differences (e.g., cross-lag analysis) rather than relying more fully on analysis of within-person change and joint analysis of individual differences in within-person change in cognition and health.

Discussion. We make the case for focusing on the interdependency between within-person changes in health and cognition and suggest methods that would support this.

IALSA: Integrative Analysis of Longitudinal Studies on Aging

- The IALSA network is currently comprised of over 30 longitudinal studies on aging, spanning eight countries.
  - Mix of population representative, volunteer, and special population samples, aged from birth to 100 years (focus 50+), with birth cohorts ranging from 1880 to 1980, assessed during historical periods from 1946 to the present. Between-occasion intervals range from 6 months to 17 years (the majority 1-5 years), with up to 32 (typically 3-5) measurement occasions spanning 4 to 48 years of within-person assessment.
- Primary goal: To facilitate new longitudinal research in ways that permit direct comparison of findings and cumulative knowledge from a within-person perspective
  - ❖ Direct involvement of PIs and research teams

Hofer & Piccinin, 2009; Piccinin & Hofer, 2008

IALSA Approach

- Coordinated analysis
  - Aim: To maximize the data value from each study while making results as comparable as possible
  - Parallel analysis models containing same covariate domains
  - Expect similar *conclusions* regardless of the exact variables used (i.e., conceptual replication; similar pattern of parameter estimates).
    - Construct-level comparison
    - Common statistical models
    - Emphasis on cross-culture, cross-study comparisons

Hofer & Piccinin, 2009, Psychological Methods

Data Comparability: Searchable Database

Longitudinal Studies on Aging

View Summary   Add/Edit a Study   Add/Edit a Measure

Studies with Common Measures/Domains:

(1) - Biomedical and Physical Functioning - Cerebrovascular / Dementia - Dementia Diagnosis
(2) - Psychosocial and Demographic - Demographics - Education
(3) - Cardiovascular / Pulmonary - (Any)
(4) - Speed - (Any)
(5) - Memory - (Any)
(6) - Stress / Life Events - (Any)

4 Studies Matching 6 Measures/Domains:	(1)	(2)	(3)	(4)	(5)	(6)
CLS	x	x	x	x	x	x
NAS	x	x	x	x	x	x
CCS	x	x	x	x	x	x
SATSA	x	x	x	x	x	x
10 Studies Matching 5 Measures/Domains:	(1)	(2)	(3)	(4)	(5)	(6)
ALSA		x	x	x	x	x
BOLSA		x	x	x	x	x
EAS	x	x	x	x	x	
GENDER		x	x	x	x	x
H-70	x	x	x	x	x	
ILSE		x	x	x	x	x
LASA		x	x	x	x	x
LSCC		x	x	x	x	x
NORA		x	x	x	x	x
OCTO-Twin	x	x	x	x	x	
6 Studies Matching 4 Measures/Domains:	(1)	(2)	(3)	(4)	(5)	(6)
HOPE	x	x	x		x	
LBLS		x	x	x	x	
SLS		x	x	x	x	
SWILSO-O		x	x	x		x
VLS		x	x	x	x	
HRS		x	x		x	x
1 Study Matching 3 Measures/Domains:	(1)	(2)	(3)	(4)	(5)	(6)
WLS		x	x			x
0 Studies Matching 2 Measures/Domains:	(1)	(2)	(3)	(4)	(5)	(6)
0 Studies Matching 1 Measure/Domain:	(1)	(2)	(3)	(4)	(5)	(6)

Return to Select Different Measures/Domains

Data Availability by Domain

- Health
  - CVD (25), BMI (24), SRHealth, diabetes (23), death (22), BP (21), stroke (18), functional (16), blood (15), dementia dx (8)
- Cognition
  - Memory, speed (20), Gf, Gc (19), attention (15), MMSE (14)
- Personality
  - Neuroticism, Extraversion (13), Openness (7)

Psychological Methods
2009, Vol. 14, No. 2, 150–164

1082-989X/09/$12.00 DOI: 10.1037/a0015566

Integrative Data Analysis Through Coordination of Measurement and Analysis Protocol Across Independent Longitudinal Studies

Scott M. Hofer and Andrea M. Piccinin
Oregon State University

Replication of research findings across independent longitudinal studies is essential for a cumulative and innovative developmental science. Meta-analysis of longitudinal studies is often limited by the amount of published information on particular research questions, the complexity of longitudinal designs and the sophistication of analyses, and practical limits on full reporting of results. In many cases, cross-study differences in sample composition and measurements impede or lessen the utility of pooled data analysis. A collaborative, coordinated analysis approach can provide a broad foundation for cumulating scientific knowledge by facilitating efficient analysis of multiple studies in ways that maximize comparability of results and permit evaluation of study differences. The goal of such an approach is to maximize opportunities for replication and extension of findings across longitudinal studies through open access to analysis scripts and output for published results, permitting modification, evaluation, and extension of alternative statistical models and application to additional data sets. Drawing on the cognitive aging literature as an example, the authors articulate some of the challenges of meta-analytic and pooled-data approaches and introduce a coordinated analysis approach as an important avenue for maximizing the comparability, replication, and extension of results from longitudinal studies.

Keywords: longitudinal, integrative data analysis, meta-analysis, data pooling, longitudinal studies

Hypothesis Testing: Cross-Cohort Comparison

- IALSA collaboration with HALCyon (UK) Network
- Joint effects of educational attainment, SES, and early life cognition on late life outcomes (adult cognition/change; mortality; dementia risk)
  - Birth cohort as natural experiment
  - e.g., Selection (SEP, childhood cognition) vs. educational attainment as protective factors
  - Measurement: Factor-level analysis of comparable cognitive constructs

Int. J. Epidemiol. Advance Access published October 28, 2012

Published by Oxford University Press on behalf of the International Epidemiological Association *International Journal of Epidemiology* 2012;1–8

doi:10.1093/ije/dys148

Benefits of educational attainment on adult fluid cognition: international evidence from three birth cohorts

Sean AP Clouston,[1,2]* Diana Kuh,[3] Pamela Herd,[4] Jane Elliott,[5] Marcus Richards[3] and Scott M Hofer[1]

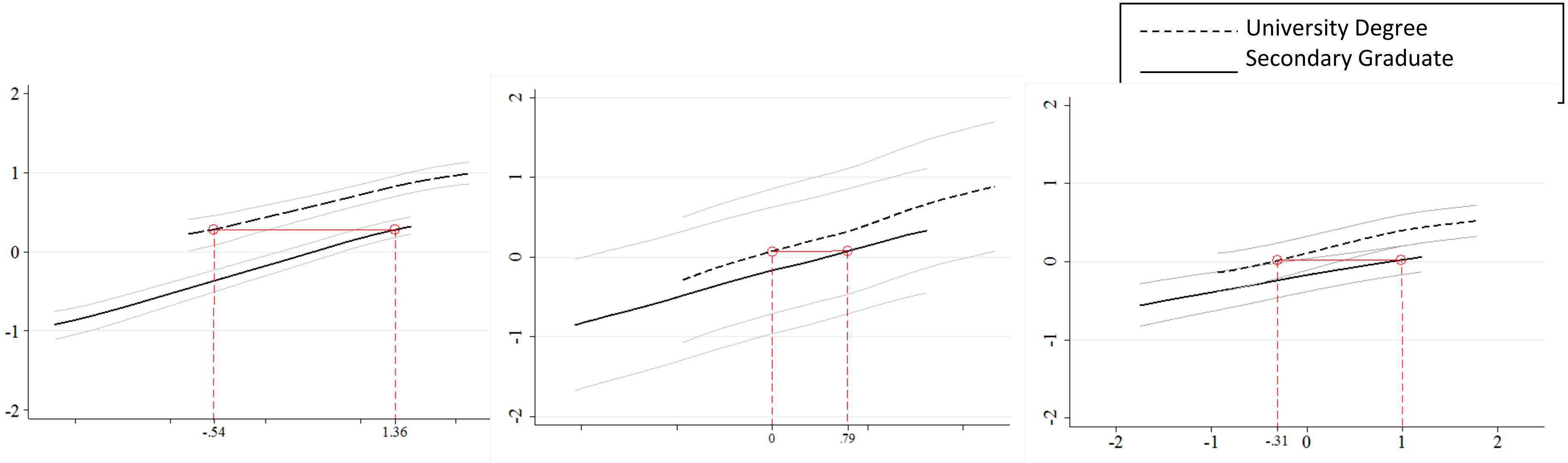
[1]Department of Psychology, University of Victoria, Victoria, BC, Canada, [2]Unit for Lifelong Health and Ageing, Medical Research Council London, UK, [3]Medical Research Council National Survey of Health and Development, Department of Epidemiology and Health Care, Royal Free and Univeresity College, London Medical School, London, United Kingdom, [4]Department of Sociology, University of Wisconsin—Madison, Madison, WI, USA and [5]Institute of Education, Centre for Longitudinal Studies, University of London, London, UK

*Corresponding author. University of Victoria, Department of Psychology, 3800 Finnerty Rd., Cornett building A236, Victoria, BC V8P 5C2, Canada. E-mail: saclous@uvic.ca

Accepted 9 August 2012

Background Educational attainment is highly correlated with social inequalities in adult cognitive health; however, the nature of this correlation is in dispute. Recently, researchers have argued that educational inequalities are an artefact of selection by individual differences in prior cognitive ability, which both drives educational attainment and tracks

Educational Benefits on Adult Cognition (Clouston et al): WLS (US 1941), NCDS (UK 1946), NSHD (UK 1958)



US 1941 WLS | UK 1946 NCDS | UK 1958 NSHD

Cognitive offset' (ΔC): Expected benefit of university level educational attainment in terms of difference in adolescent cognition required to reach same level of adult cognition

Adult cognition was indicated by WAIS similarities (WLS) and indicators of verbal memory, verbal learning, and category fluency (NCDS, NSHD)

Coordinated Analysis

- Brown, C.L., Gibbons, L.E., Kennison, R.F., Robitaille, A., Lindwall, M., Mitchell, M., Shirk, S.D., Atri, A., Cimino, C.R., Benitez, A., MacDonald, S.W.S., Zelinski, E., Willis, S.L., Schaie, K.W., Johansson, B., Dixon, R.A., Mungas, D.M., Hofer, S.M. & Piccinin, A.M. (2012). **Social activity and cognitive functioning over time.** *Journal of Aging Research*, vol. 2012, Article ID 287438.
- Lindwall, M., Cimino, C. R., Gibbons, L. E., Mitchell, M., Benitez, A., Brown, C. L., Kennison, R. F., Shirk, S. D., Atri, A., Robitaille, A., MacDonald, S. W. S., Zelinski, E., Willis, S. L., Schaie, K. W., Johansson, B., Praetorius, M., Dixon, R. A., Mungas, D. M., Hofer, S. M. & Piccinin, A. M. (2012). **Dynamic associations of change in physical activity and change in cognitive function: Coordinated analyses of four longitudinal studies.** *Journal of Aging Research*, Article ID 493598.
- Mitchell, M., Cimino, C. R., Benitez, A., Brown, C. L., Gibbons, L. E., Kennison, R. F., Shirk, S. D., Atri, A., Robitaille, A., Lindwall, M., MacDonald, S. W. S., Zelinski, E., Willis, S. L., Schaie, K. W., Johansson, B., Dixon, R. A., Mungas, D. M., Hofer, S. M. & Piccinin, A. M. (2012). **Cognitively stimulating activities: Effects on cognition across four studies with up to 21 years of longitudinal data.** *Journal of Aging Research*. Article ID 461592.

Coordinated Analysis

- Clouston, S., Brewster, P., Kuh, D., Richards, M., Cooper, R., Hardy, R., Rubin, M., & Hofer, S. M. (in press). **The dynamic relationship between physical function and cognition in longitudinal aging cohorts: A systematic review.** *Epidemiological Reviews*.
- Clouston, S., Kuh, D., Herd, P., Elliott, J., & Richards, M., & Hofer, S. M. (in press). **Benefits of educational attainment on adult fluid cognition: International evidence from three birth cohorts.** *International Journal of Epidemiology*.
- Piccinin, A. M., Muniz, G., Clouston, S.A., Reynolds, C.A., Thorvaldsson, V.Deary, I., Deeg, D., Johansson, B., MacKinnon, A., Spiro, A. III, Starr, J. M., Skoog, I. & Hofer, S. M. (In press). **Integrative analysis of longitudinal studies on aging: Coordinated Analysis of age, sex, and education effects on change in MMSE scores.** Journal of Gerontology: Psychological Sciences.

Main Points

- Cross-study comparison can proceed without quantitative harmonization
- Quantitative comparison across studies requires better understanding of properties of particular measures
  - Scale-level and item-level analysis
- Next steps (CLOSER; IALSA Program Project Grant)
  - Harmonization platform including metadata catalogue and harmonization rules (logical, psychometric)
    - Develop "Longitudinal Studies of Aging Data Schemas" that will facilitate harmonization across studies.
  - New item-level data collection for psychometric harmonization
    - Identify item-level data required for co-calibration
    - Item library for quantitative comparison across past, current, and future studies

Challenges for Harmonization: Longitudinal Studies of Aging

- What to do when measurements differ?
  - Measurement calibration (e.g., Lowest common denominator)
  - Selection of most comparable studies
  - Comparison at construct level
- When possible, calibration isolates measurement differences, but other differences remain
  - Birth cohort, sample (e.g., country/culture; representativeness), attrition/mortality, period effects
  - Important for evaluating generalizability

Are differences in study results due to measurement differences?

- Sensitivity to measurement properties
  - Reliability
  - Floor/Ceiling Effects
  - Distribution of item difficulty (e.g., sensitive at different levels of ability)
  - Measures different primary processes of higher-order construct
  - Language
  - Variations in administration/scoring

Crystallized Knowledge

	WAIS Comprehension	WAIS Information	WAIS Similarities	Synonyms	Vocabulary	NART
ALSA			X		X	X
BAS		X	X		X	
BOLSA	X	X	X	X		
CCS			X			X
CHS						
CLS			X		X	X
EAS		X	X		X	
GENDER				X		
H-70				X		
HOPE						X
HRS			X		X	
ILSE		X	X			
LASA						
LBLS					X	
LSCC			X		X	
NAS					X	
NORA						
NSHD						X
OBAS	X		X		X	
OCTO-T		X		X		
SATSA		X	X	X		
SLS				X		
SWILSO						
UNCAHS						
VLS	X	X			X	
WLS			X			

Fluid Reasoning

	Arithmetic	WAIS Object Assembly	WAIS Picture Arrangement	WAIS Picture Completion	WAIS Block Design	Series	Matrices	Rotation	General Aptitude Test Battery	Concept Formation
ALSA										
BAS		X			X		X			
BOLSA	X	X	X	X	X		X			
CCS							X			
CHS										
CLS										
EAS					X					
GENDER					X		X			
H-70					X		X			
HOPE							X			
HRS	X									
ILSE				X	X					
LASA							X			
LBLS						X		X		X
LSCC							X			
NAS								X	X	
NORA							X			
NSHD										
OBAS				X	X		X			
OCTO-T					X		X			
SATSA							X	X		
SLS	X					X		X		
SWILSO										
UNCAHS										
VLS						X		X		
WLS									X	

Speeded Performance

	Figure Identification	Substitution Coding Tasks	Number Copy Task	Identical Pictures	Number Comparison	Finding As, Is and Os	Reaction Time
ALSA		X					
BAS		X	X				
BOLSA		X					X
CCS							X
CHS		X					
CLS		X					X
EAS		X	X				X
GENDER	X						
H-70	X						
HOPE							
HRS							
ILSE		X					
LASA		X					
LBLS	X				X		
LSCC		X				X	
NAS				X			
NORA		X					X
NSHD						X	
OBAS		X				X	X
OCTO-T	X	X					
SATSA	X	X					
SLS				X	X	X	
SWILSO						X	
UNCAHS							
VLS		X	X	X	X		X
WLS							

Memory (episodic)

	Category Cued Recall	Selective Reminding	Prose Recall	Word List Free Recall	Delayed Word List Free Recall	MIR Memory Test	Incidental Memory	Coin Test	Everyday Memory
ALSA	X			X			X		
BAS		X							
BOLSA									
CCS			X				X		X
CHS									
CLS									
EAS		X	X	X					
GENDER				X	X				
H-70			X	X	X	X		X	
HOPE			X						
HRS				X	X		X		X
ILSE				X	X				
LASA				X					X
LBLS	X		X	X	X				
LSCC				X	X		X		
NAS				X	X				
NORA			X						
NSHD			X	X	X				X
OBAS			X	X	X				
OCTO-T			X			X		X	
SATSA			X	X	X	X		X	
SLS				X	X				
SWILSO									
UNCAHS									
VLS	X		X	X					
WLS									

IALSA

Piccinin, A.M., Muniz, G., Clouston, S., Reynolds, C.A., Thorvaldsson, V., Deary, I.J., Deeg, D., Johansson, B., Mackinnon, A., Spiro, A., Starr, J.M., Skoog, I., Hofer, S.M., (2012). Coordinated analysis of age, sex, and education effects on change in MMSE scores. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, doi:10.1093/geronb/gbs077

Coordinated Analysis of Age, Sex, and Education Effects on Change in MMSE Scores

Andrea M. Piccinin,[1] Graciela Muniz-Terrera,[2] Sean Clouston,[1] Chandra A. Reynolds,[3] Valgeir Thorvaldsson,[4] Ian J. Deary,[5] Dorly J.H. Deeg,[6] Boo Johansson,[4] Andrew Mackinnon,[7] Avron Spiro, III,[8] John M. Starr,[9] Ingmar Skoog,[10] and Scott M. Hofer[1]

[1]Department of Psychology, University of Victoria, British Columbia, Canada.
[2]MRC Unit for Lifelong Health and Ageing, London.
[3]Department of Psychology, University of California, Riverside.
[4]Department of Psychology, University of Gothenburg, Sweden.
[5]Department of Psychology, University of Edinburgh.
[6]EMGO Institute for Health and Care Research, VU University Medical Centre/LASA, Amsterdam, The Netherlands.
[7]Centre for Youth Mental Health Research, University of Melbourne, Australia.
[8]Normative Aging Study, VA Boston Healthcare System, Boston, MA.
[9]Alzheimer Scotland Dementia Research Centre, University of Edinburgh.
[10]Neuropsychiatric Epidemiology, University of Gothenburg, Sweden.

Objectives. We describe and compare the expected performance trajectories of older adults on the Mini-Mental Status Examination (MMSE) across six independent studies from four countries in the context of a collaborative network of longitudinal studies of aging. A coordinated analysis approach is used to compare patterns of change conditional on sample composition differences related to age, sex, and education. Such coordination accelerates evaluation of particular hypotheses. In particular, we focus on the effect of educational attainment on cognitive decline.

Method. Regular and Tobit mixed models were fit to MMSE scores from each study separately. The effects of age, sex, and education were examined based on more than one centering point.

Results. Findings were relatively consistent across studies. On average, MMSE scores were lower for older individuals and declined over time. Education predicted MMSE score, but, with two exceptions, was not associated with decline in MMSE over time.

Conclusion. A straightforward association between educational attainment and rate of cognitive decline was not supported. Thoughtful consideration is needed when synthesizing evidence across studies, as methodologies adopted and sample characteristics, such as educational attainment, invariably differ.

Key Words: Cognitive— Coordinated Analysis— Education—Longitudinal—Mental Status Exam— Meta-analysis—Mixed Model.

Education and MMSE

- Consistent differences in performance
  - Concern regarding bias
- Inconsistent findings regarding rate of change
  - Higher education, less decline:
    - YES: Change score analyses; Jaqmin-Gadda, 1997
    - NO: Growth curve analyses
    - YES and NO:
      - Change point analysis: Hall, 2007;
      - Linear v Quadratic growth models with education squared; Wilson, 2009
      - Growth mixture-logistic survival analysis: Muniz et al, 2010

Analysis

- Compare results based on coordinated analysis
  - Consider impact of covariate centring
- Describe impact of education on cognitive function
  - "Measurement" of educational attainment
- Address MMSE ceiling (Tobit model)
- Between person age differences versus within person age changes

Participating Studies

Name	Start Yr	n (T1)	Age (T1)	Followup (yrs)	Occ Interval	Curr # Occ	Type Sample
Swedish Adoption Twin Study of Aging (SATSA)	1984	1500	40-84	12	3	5	Swedish Twin Registry
Longitudinal Aging Study Amsterdam (LASA)	1991	3107	55-85	12	3	5	Urban and rural municipal registries
Healthy Older Person Edinburgh Study (HOPE)	1990	603	70+	12	4	4	Medical registry
Canberra Longitudinal Study (CLS)	1991	897	70-93	14	3.5	5	Community sample (electoral role), institutional care, oversampling of very old
Origins of Variance in the Old-Old: Octogenarian Twins (OCTO-Twin)	1990	702	80+	8	2	5	Swedish Twin Registry
Gerontological and Geriatric Population Studies in Gothenburg, Sweden (H-70)	1971 1986	1000 396	70 85	29 14	2-5 2-4	12 6	Representative sample: Gothenburg; sequential design

Administration Variations

Original MMSE Items	SATSA	LASA	HOPE	CLS	OCTO-Twin & H-70
STATE	**Country/Land**	Province	**Country**	State	**Country**
COUNTY	County	**Address**	County	**Country**	County
CITY/TOWN	City/Town	Municipality	City	Town	**Place**
HOSPITAL	**District (municipal)/Institution**	**Two main streets in neighbourhood**	**Residence**	**Residence**	**District/Institution**
FLOOR OF BLDG	**Address/Department**	Floor of building	Floor	Floor	**Street/ward**
APPLE TABLE PENNY	**nykel, tandborste, lampa (Key, toothbrush, lamp)**	**Appel Tafel Stuiver**	**Lemon key ball**	Apple table penny	**Key, toothbrush, lamp**
SERIAL 7s	Serial 7s	Serial 7s	**WORLD backward**	Serial 7s	Serial 7s
		(alt) DORST backward	**(alt) Serial 7s**	(alt) WORLD backward	
PENCIL IDENTIFIED	**pen**	pencil	pencil	pencil	pencil
'NO IFS....' REPEATED	**"burned down two-family house"**	"No ifs..." ("Geen als en of maar")	"No ifs…" repeated	"No ifs…" repeated	**"burned down two-family house"**
Right Hand	**Hand**	Right Hand	Right hand	Right hand	**Hand**
Put it on the floor	Put it on the floor	Put it on **your lap**	Put it on the floor	Put it on **your lap**	Put it on the floor
CLOSED EYES	**Point at the door**	Closed eyes	Closed eyes	Closed eyes	**Point at the door**
Additional scoring details:		***Best of*** Serial 7s / World Backward Two versions of memory test to reduce practice effect		***Best of*** Serial 7s/ World Backward	
Language of administration:	Swedish	Dutch	English	English	Swedish

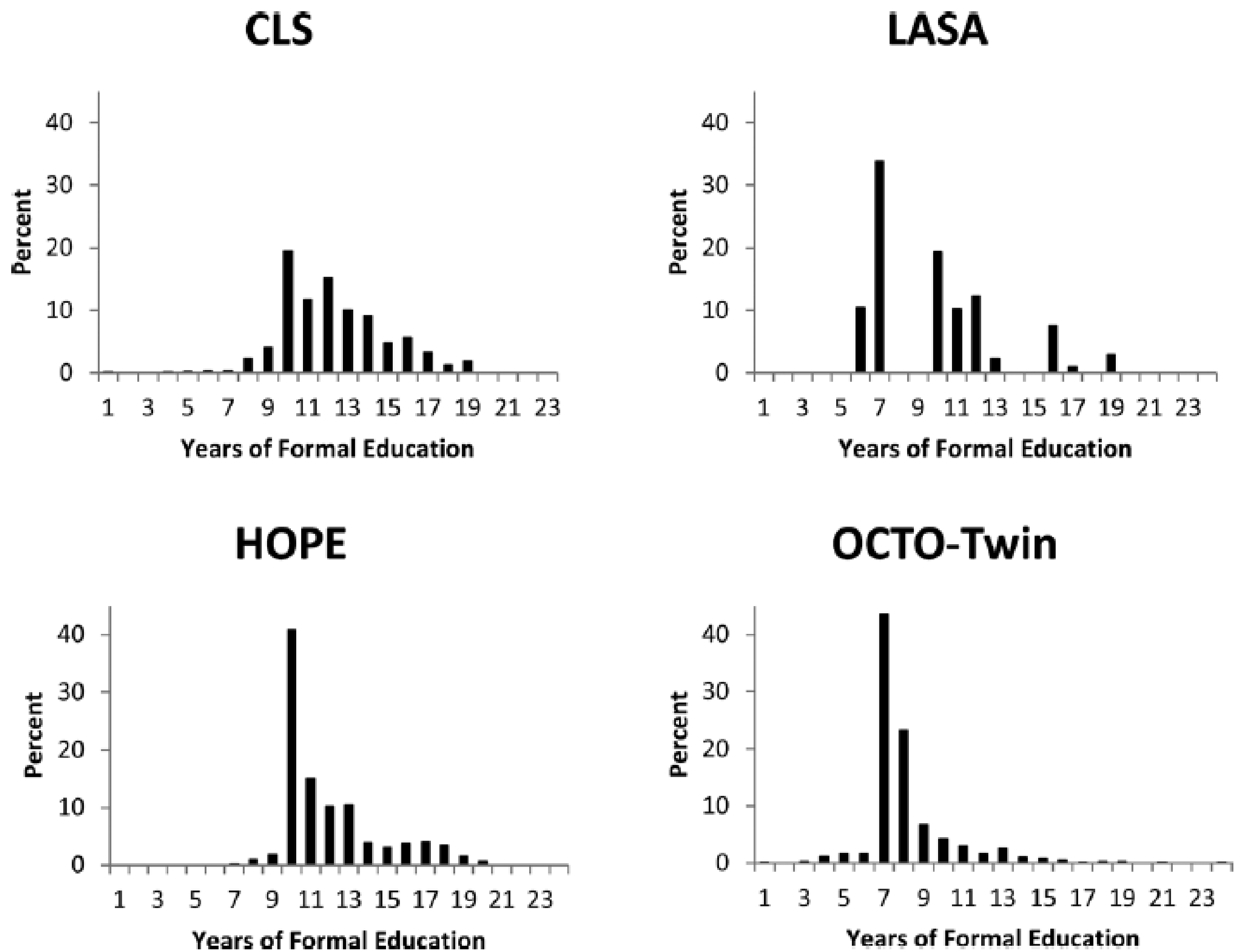
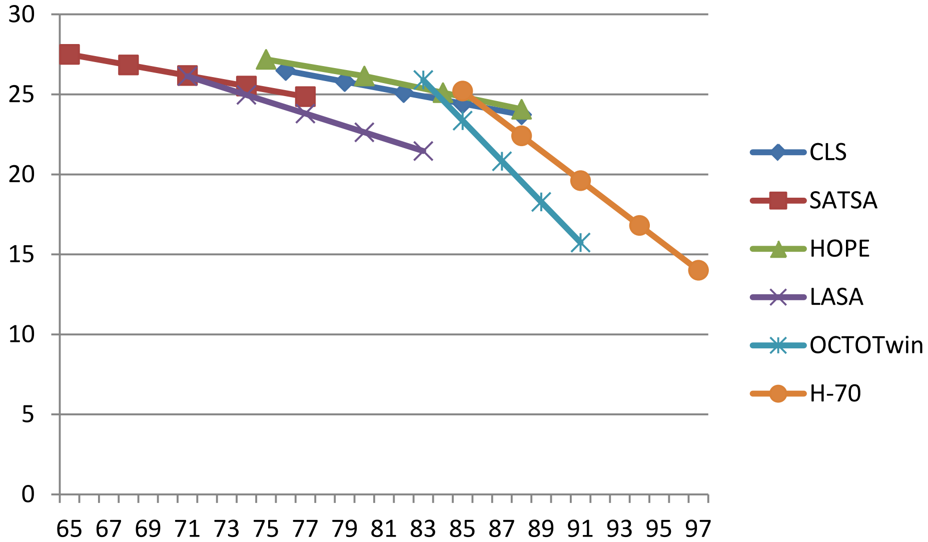
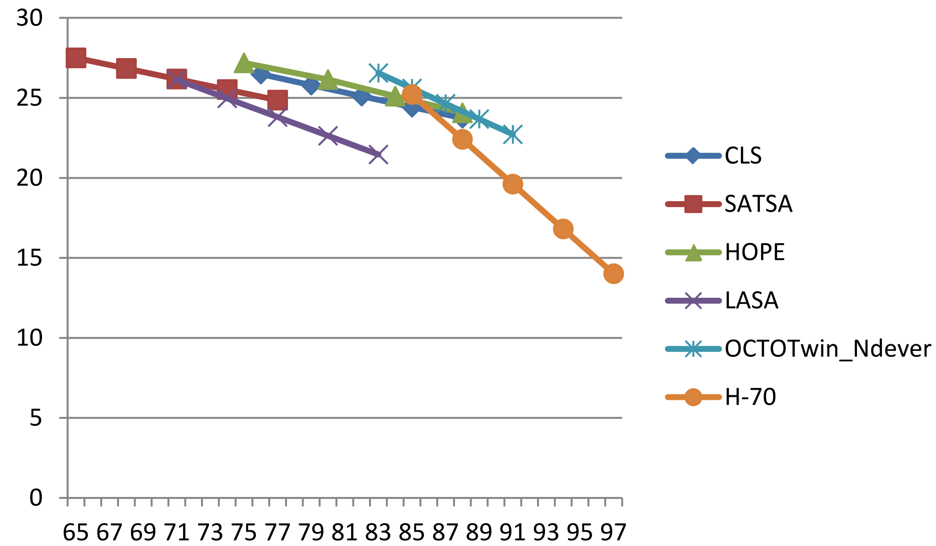


Figure 1. Education distribution (%) by study. *Note*. Gerontological and Geriatric Population Studies in Gothenburg, Sweden [H-70] and Swedish Adoption/Twin Study of Aging [SATSA]: 60% completed elementary school or less and 40% completed more than elementary school. CLS = Canberra Longitudinal Study; HOPE = Healthy Older Person Edinburgh; LASA = Longitudinal Aging Study Amsterdam; OCTO-Twin = Origins of Variance in the Oldest-Old.

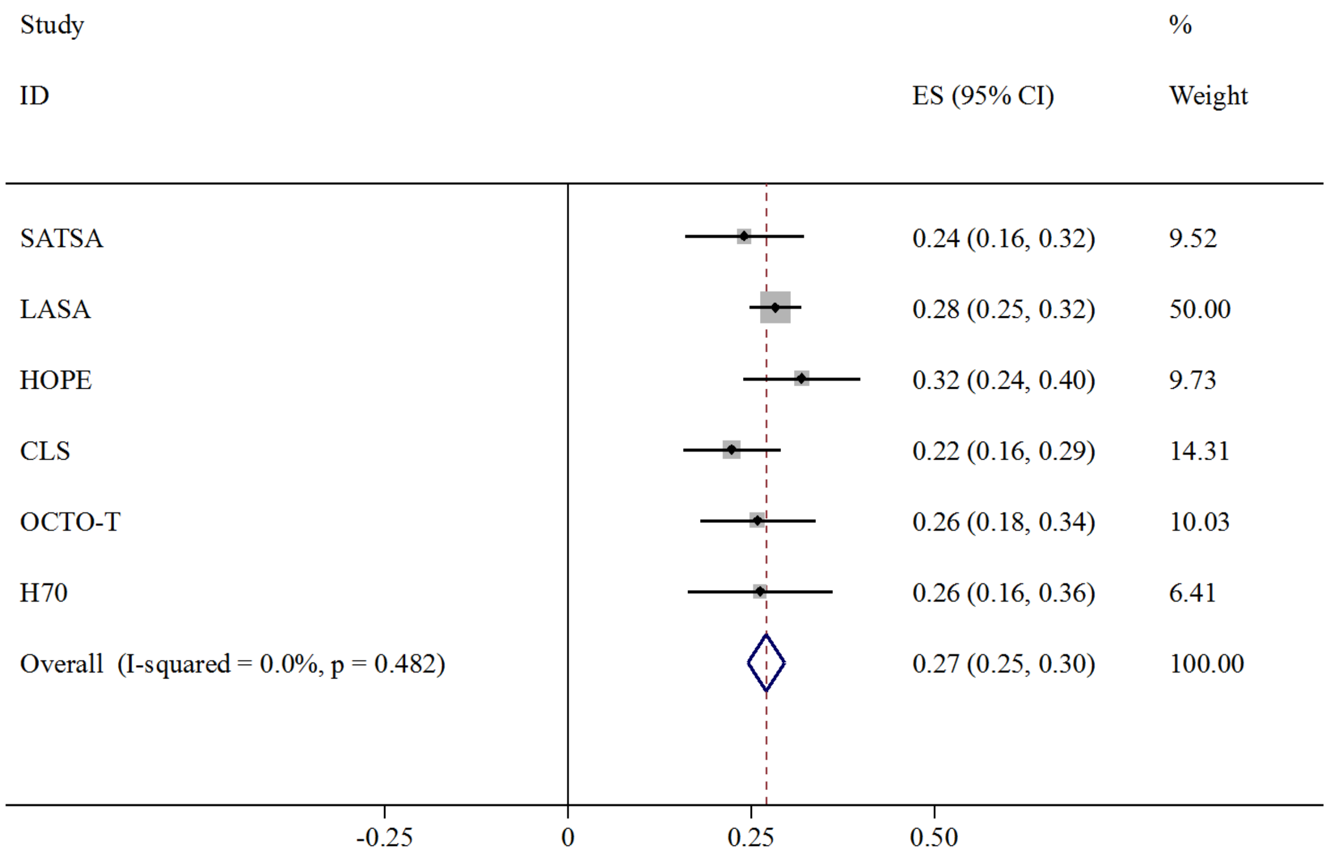
Predicted Trajectories



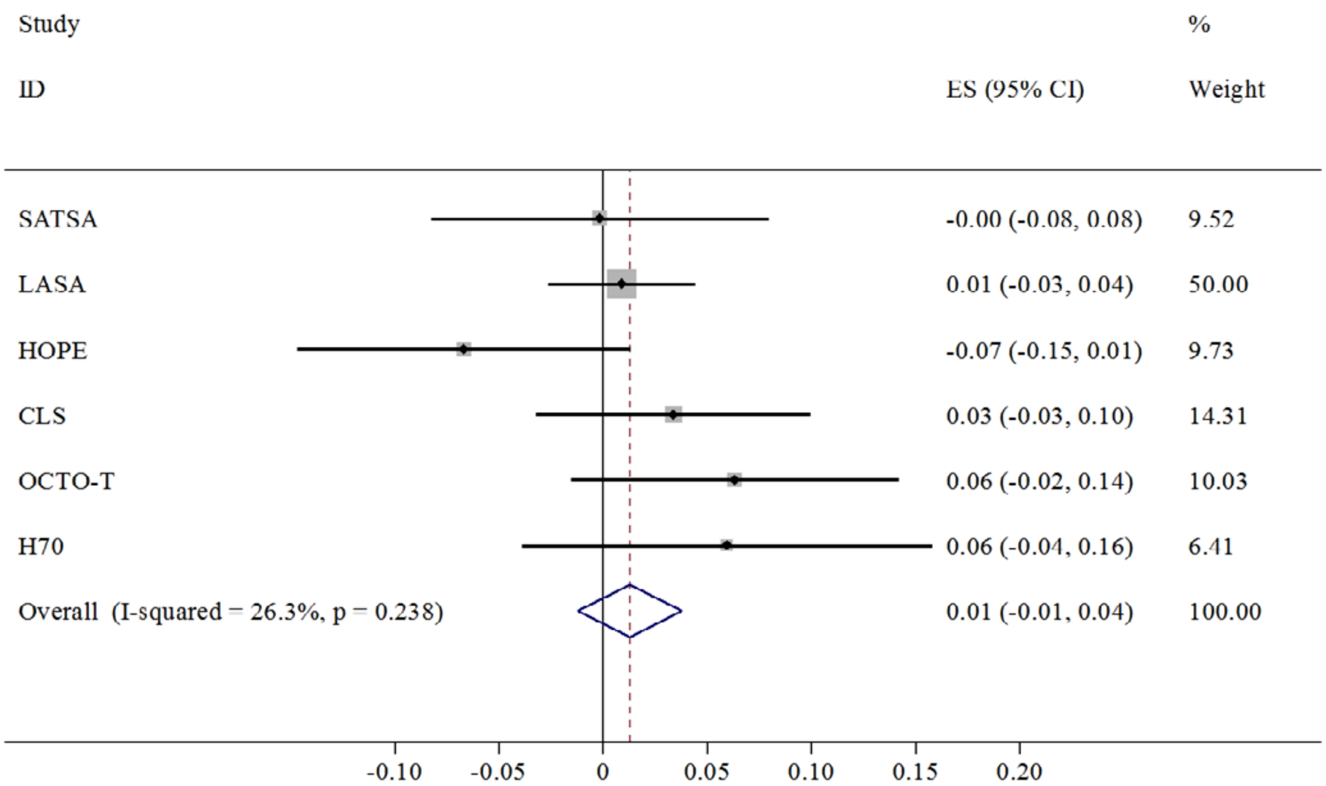
Hypothetical Individuals
- male
- at study median age and education
- followed for length of study



Meta-analytic Results: Educational Attainment Intercepts



Education * Time



NB: Estimates have been standardized to account for sample size heterogeneity.
OCTO-Twin study: non-demented estimates for change in educational attainment.

Simultaneous Replication with Different Measures

- Pre-standardization using relative age and education differences
  - Between-Person Mean and SD for relative age and education group in common across studies at T1
  - Individuals aged 80-85 with 6-10 years education for each sample.

Sub-sample Characteristics

Study	N	Age	Median Education	# waves	Inter-wave	Follow-up	Sampling
HOPE	596	70+	10	3	4	8	Local medical registers; strict health criteria
LASA	2991	55-85	9	5	3	12	Municipal registries
OCTO	290	80+	7	5	2	8	National twin registry

Selection for analysis:

- MMSE >19 at wave 1

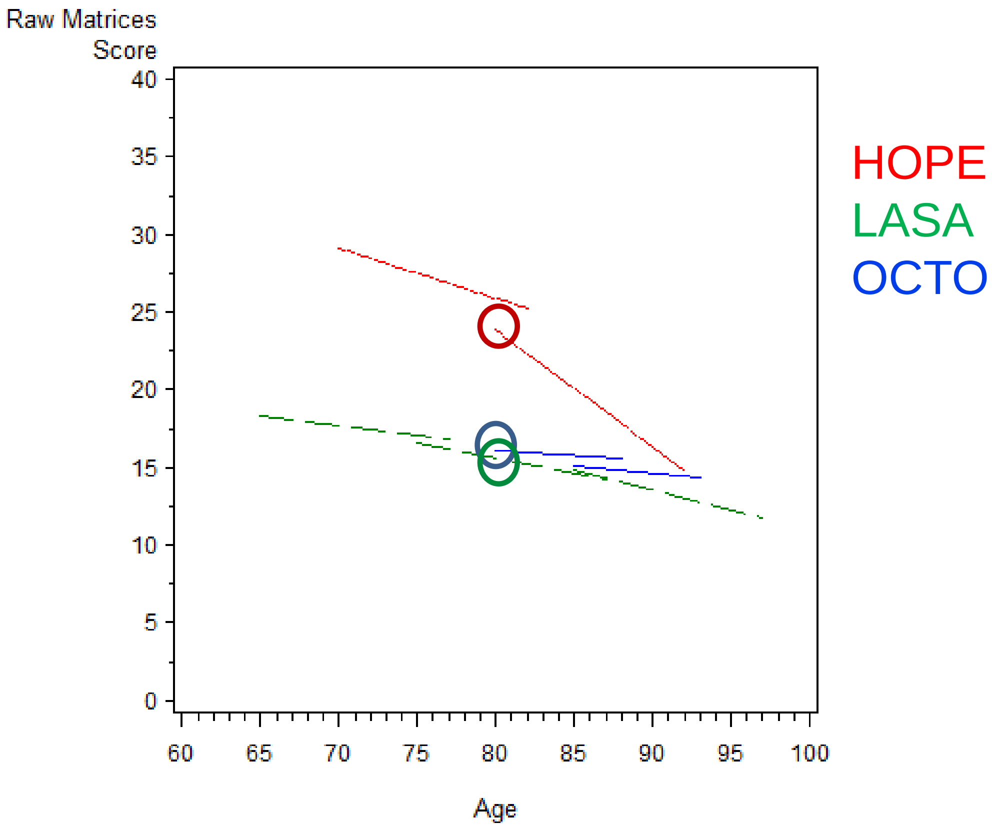
Measure Characteristics

Study	Matrices	Episodic Memory	Speed	Vocabulary
HOPE	Ravens Standard	WMS-R Logical Mem	--	NART
LASA	Ravens Coloured	Rey-AVLT	Alphabet code	--
OCTO	Figure Logic	Prose recall	Symbol Digit	Synonyms

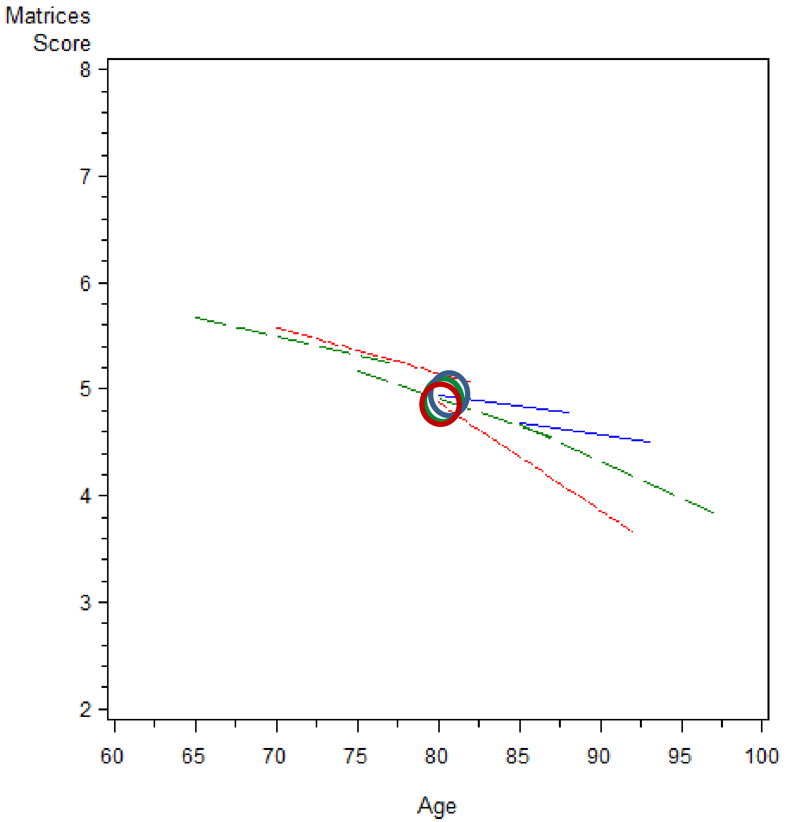
IALSA

Model-predicted Trajectories: Matrices

Raw Score



Relative Age & Education Standardized



Challenges for Harmonization: Longitudinal Studies of Aging

- What to do when measurements differ?
  - Select items that are congruent (lowest common denominator)
  - Selection of studies with identical measures
  - Compare standardized effects (regression standardization)
  - Construct level comparison
- Replication and generalizability of key results, hypotheses, and theories within sociohistorical context
  - Birth cohort and country differences as natural experiment: Evaluation of joint effects of educational attainment, parental SES, and early life cognition on late life outcomes (cognitive level/change; mortality; dementia risk)

Approaches to Psychometric Harmonization

- Item vs. Scale level analysis
- Reflective measures (item level analysis)
  - Personality, Gf, Gc, Gq
  - Gsm: Memory (sensitivity to item context)
- Formative and mixed measures
  - Depressive symptoms
  - Diagnostic checklists
- Measures without items (test level analysis)
  - Processing speed
  - Verbal fluency

No Overlap

Sample#1	Measure#1		
Sample#2	1-2	Measure#2	
Sample#3	1-3	2-3	Measure#3

- No sample with any of 3 covariances (1-2, 1-2, 2-3)
- Use "Conceptual Replication" framework
- Collect additional data where needed covariances are observed
  - Consider existing sample characteristics when drawing additional sample

Bontempo. D. E., Hofer, S. M. & Piccinin, A. M. (2009, August). Factorial invariance and scale comparison across studies when measures differ. In S. M. Hofer and K. Grimm (Chairs), *Evaluating and optimizing measurement comparison across longitudinal studies*. Paper symposium conducted at the annual meeting of the American Psychological Association, Toronto.

Psychometric Harmonization: Factor Analytic "Scaffolding"

- Extension analysis can be used when a study has multiple indicators of a particular domain
  - Increased probability of overlapping measures
    - e.g., Study 1 has measures A, B, C
    - Study 2 has measures A, D, E
    - Study 3 has measures B, C, F
  - Most common in cognitive research
    - e.g., crystallized ability indicated by vocabulary, information, similarities

Overlap: Common Metric via CFA

Study	Measure #1	Measure #2	Measure #3	Measure #4
S#0	x			
S#1	x	x		
S#2		x	x	
S#3		x		
S#4	x	x		
S#5			x	x
S#6		x		
S#7				x
S#8		x	x	
S#9	x			

Exploit observed measure covariances to place measures on a common factor scale using CFA model

Where multiple group CFA models are possible, each model group must have coverage of each covariance

- Using multiple groups can mitigate group mean and variance differences from biasing common factors by permitting groups to differ in factor mean and variance
- Multiple groups would permit cross-group factorial invariance tests
- In single-group models it may still be useful to employ covariates (e.g., MIMIC) to model intercept non-invariance; but loading non-invariance cannot be addressed

- This approach could be employed in 3 possible subsets, but not across all samples
- The 1st two subsets could support 2-group CFA models

Study	Measure #1	Measure #2
S#0	x	
S#1	x	x
S#2		x
S#3		x
S#4	x	x
S#6		x
S#8		x
S#9	x	

Study	Measure #2	Measure #3
S#1	x	
S#2	x	x
S#3	x	
S#4	x	
S#6	x	
S#8	x	x

Study	Measure #3	Measure #4
S#2	x	
S#5	x	x
S#7		x
S#8	x	

Common Metric via CFA models

Graham, J. W., Hofer, S. M., & MacKinnon, D. P. (1996). Maximizing the usefulness of data obtained with planned missing value patterns: An application of maximum likelihood procedures. *Multivariate Behavioral Research, 31(2), 197-218.*

McArdle, J. J. (1994). Structural factor analysis experiments with incomplete data. *Multivariate Behavioral Research, 29(4), 409-454.*

- Follows the logic of planned missing value patterns
  - When missing patterns are across sample versus waves consideration should be given to sampling heterogeneity
- Depending on covariance coverage, it is likely that not all cases can be used
- Like the “No Overlap” situation, additional data can be collected to address incomplete coverage
- Generally applied with scale scores

Common Metric via CFA models

- Can place ***scale scores*** on a common factor or utilize item-level data to place ***factor scores*** onto a higher order common factor.
  - Item-level models do permit investigations of non-invariance at much greater resolution, and/or use of an invariant subset of scale items
  - Some correspondence of item-level CFA harmonization and IRT harmonization

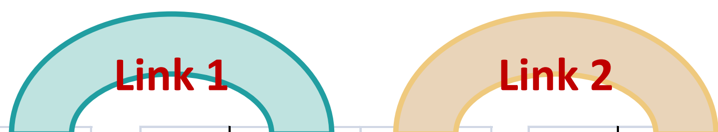
Curran, P. J., Hussong, A. M., Cai, L., Huang, W., Chassin, L., Sher, K. J., et al. (2008). Pooling data from multiple longitudinal studies: The role of item response theory in integrative data analysis. *Developmental Psychology, 44(2), 365-380.*

McArdle, J. J., Grimm, K. J., Hamagami, F., Bowles, R. P., & Meredith, W. (2009). Modeling life-span growth curves of cognition using longitudinal data with multiple samples and changing scales of measurement. *Psychological Methods, 14(2), 126-149.*

- CFA methods may be more flexible because they permit some group differences to be modeled (or use of MIMIC)

Multistep CFA Harmonization

- Can we use a multi-step procedure to link our 3 subsets, and thus use all cases?
  - A necessary condition would require an observed covariance to link subsets
  - Other requirements and conditions are the subject of ongoing research efforts



Study	Measure #1	Measure #2
S#0	x	
S#1	x	x
S#2		x
S#3		x
S#4	x	x
S#6		x
S#8		x
S#9	x	

Study	Measure #2	Measure #3
S#1	x	
S#2	x	x
S#3	x	
S#4	x	
S#6	x	
S#8	x	x

Study	Measure #3	Measure #4
S#2	x	
S#5	x	x
S#7		x
S#8	x	

Implementing The Steps

- Step#1: obtain parameters for Measure#2 and Measure#3 using the 2nd subset, outputting factor scores.
- Step#2: obtain parameters for Measure#1 in the 1st subset, fixing the parameters for Measure#2 to those obtained in Step#1. Output factor scores.
- Step#3: obtain parameters for Measure#4 in the 3rd subset, now fixing the parameters for Measure#3 to those obtained in Step#1. Output factor scores.

Link 1

Link 2

Study	Measure #1	Measure #2
S#0	x	
S#1	x	x
S#2		x
S#3		x
S#4	x	x
S#6		x
S#8		x
S#9	x	

Study	Measure #2	Measure #3
S#1	x	
S#2	x	x
S#3	x	
S#4	x	
S#6	x	
S#8	x	x

Study	Measure #3	Measure #4
S#2	x	
S#5	x	x
S#7		x
S#8	x	

Theoretical Framework for Multistep Harmonization

- Linking across steps is accomplished by fixing parameters for the linking measure in the subsequent step to parameters obtained for that measure in the initial step
- The fixed parameters are expected to lock the factor in conceptual space, and the estimated item loadings for the extension instrument are constrained to load as best as they can on this “locked down” factor
- Elements of this reasoning draw on both Extension Analysis and Partial Measurement Invariance

Extension Analysis

- Extension analysis utilizes the loadings from a *core set* of indicators obtained from an EFA solution, and the inter-correlations of the core indicators with a set of ***extension indicators***, to derive loadings (on the previously estimated common factor) for the extension indicators.
- The extension indicators do not get to define the factor, nor do they contribute to the factor variance; but subsequent factor scoring procedures can utilize the extension loadings.
- Conceptually this has strong parallels to a multistep procedure in which freely estimates loadings for ***new items*** are obtained while the factor solution is "locked down" by the ***fixed items***
  - But new items do contribute to common factor variance
- Some cautions:
  - Extension analysis is an EFA technique and has not been studied in CFA
    - It is unclear if new item loadings would be sensitive to spurious correlations between fixed and new items. Excluding any spurious correlation is a prominent issue in Extension Analysis
  - Most work on extension analysis is silent regarding mean structures

Dwyer, P. S. (1937). The determination of the factor loadings of a given test from the known factor loadings of other tests. *Psychometrika, 2, 173-178.*

Horn, J. L. (1973). On extension analysis and its relation to correlations between variables and factor scores. *Multivariate Behavioral Research, 8(4), 477-489.*

PMI – Partial Measurement Invariance

- A multi-group CFA technique which holds that a critical mass of factorially-invariant indicators can sufficiently "lock down" a factor in conceptual space to permit a finite number of non-invariant indicators to be modeled with freely estimated factor parameters.
    - Has been invoked for groups within a sample, but also across samples in multi-national research.
- The parameters of invariant items (i.e., loadings, intercepts) are constrained to equivalence across groups while parameters for non-invariant indicators are estimated in each group.
- Caution: while intuitive, the mechanism of PMI has not been mathematically proven or subjected to rigorous empirical study.

Illustration: Depressive Symptoms

Study	CESD	SF	GADS	PAS	N
S#0	CESD-20				2087
S#1	CESD-10	SF-36			13716
S#2		SF-36	GADS		12432
S#3		SF-36			7296
S#4	CESD-10	SF-36			3654
S#5			GADS	PAS	1135
S#6		SF-36			6164
S#7				PAS	1000
S#8		SF-12	GADS		2551
S#9	CESD-20				630

Use of item-level data permits pre-harmonization of long forms into short forms

Many dichotomous items require use of polytomous factor models

1. Obtain SF-12 parameters using the subset comprising S#1, S#2, S#3, S#4, S#6, S#8. (Total n = 45813). Output factor scores.
2. Fixing SF-12 item measurement parameters (loadings, thresholds) to step#1 values, obtain measurement parameters for CESD-10 in SF-12 metric using the subset comprising S#0, S#1, S#2, S#3, S#4, S#6, S#8. (Total n = 48530). Output factor scores.
3. Fixing SF-12 item measurement parameters to step#1 values, obtain measurement parameters for GADS items in SF-12 metric using the subset comprising S#1, S#2, S#3, S#4, S#5, S#6, S#8. (Total n = 46948). Output factor scores.
4. Now fixing GADS item measurement parameters to step#3 values, obtain measurement parameters for PAS items in SF-12 metric (via GADS in SF-12) using the subset comprising S#2, S#5, S#7. (Total n = 15851). Output factor scores.
5. Merge factor scores from each step.
   – Some cases will have a factor score from multiple step. Take mean, or use a selection rule.

Bontempo. D. E., **Hofer, S. M.** & Piccinin, A. M. (2009, August). Factorial invariance and scale comparison across studies when measures differ. In S. M. Hofer and K. Grimm (Chairs), *Evaluating and optimizing measurement comparison across longitudinal studies*. Paper symposium conducted at the annual meeting of the American Psychological Association, Toronto.

Factor Score Inter-Correlations

H_SF_CD	SF12	CESD
SF12 Only	1.0000 24962	
CESD Only		0.9997 2690
SF12 or CESD	0.9750 38122	0.9497 1580
SF12 and CESD	0.9289 13160	0.9412 13160

Native	sf12	cesd	gads	pas
sf12	1			
cesd	0.76	1		
gads	0.69	.	1	
pas	.	.	0.66	1

- Correlations >.92 are encouraging
- The perfect correlation of SF12 factor scores with harmonized scores on SF12 metric is encouraging
- The High correlation of CESD factor scores with harmonized scores is surprising, given the .76 correlation of SF12 and CESD scores
- The higher correlation with CESD for cases having both SF12 and CESD scores suggests that the factor score algorithm may favor the scale with more questions, or otherwise needs more careful examination

Sensitivity to Decisions

- Starting Point – e.g., SF12 or SF12-CESD
- "Locked" Parameters – loadings and thresholds or just loadings? Consequences?
- Multi-group – would using as many groups as possible in step#1 provide a set of fixed-loadings less tainted by cross-sample differences in factor variance?
- Standardizing Variance – if factor variance was estimated at each step, would this variance be in units of the factor variance in the model that supplied the fixed loadings?
- Sample Characteristics – would an age band seen in all studies have been best?
  - Do mean differences between samples bias the factor?
  - What cross-sample differences on the harmonized factor are valid for different contexts of calibration?

Factor Models

- The ability to identify and estimate these multi-step models, and their overall good fit, is encouraging.
- Further work investigating alternative decisions, and examining the influence of factor scoring procedures should be undertaken.
- Identifying conditions where this approach works best would be a useful guide for other researchers.

IALSA

Psychological Methods
2009, Vol. 14, No. 2, 126–149

1082-989X/09/$12.00 DOI: 10.1037/a0015857

Modeling Life-Span Growth Curves of Cognition Using Longitudinal Data With Multiple Samples and Changing Scales of Measurement

John J. McArdle
University of Southern California

Kevin J. Grimm
University of California, Davis

Fumiaki Hamagami
University of Virginia

Ryan P. Bowles
Michigan State University

William Meredith
University of California, Berkeley

The authors use multiple-sample longitudinal data from different test batteries to examine propositions about changes in constructs over the life span. The data come from 3 classic studies on intellectual abilities in which, in combination, 441 persons were repeatedly measured as many as 16 times over 70 years. They measured cognitive constructs of vocabulary and memory using 8 age-appropriate intelligence test batteries and explore possible linkage of these scales using item response theory (IRT). They simultaneously estimated the parameters of both IRT and latent curve models based on a joint model likelihood approach (i.e., NLMIXED and WINBUGS). They included group differences in the model to examine potential interindividual differences in levels and change. The resulting longitudinal invariant Rasch test analyses lead to a few new methodological suggestions for dealing with repeated constructs based on changing measurements in developmental studies.

University of Victoria

Table 1
Summary of Available Data From Multiple Testing Occasions for Three Longitudinal Studies

Age (years)	Berkeley Growth (n = 61)	Guidance–Control (n = 206)	Bradway–McArdle (n = 111)
2–5½			SB-L, SB-M (111)
6	1916 SB (60)	1916 SB (205)	
7	1916 SB (47), SB-L (8)	1916 SB (204)	
8	SB-L (51)	SB-L (187)	
9	SB-L (53)	SB-L (94), SB-M (98)	
10	SB-M (53)	SB-L (102), SB-M (88)	
11	SB-L (48)	SB-L (77)	
12	SB-M (50)	SB-L (90), SB-M (43)	
13–14	SB-L (42)	SB-L (82), SB-M (97)	SB-L (111)
15		SB-M (51)	
16	WB-I (48)		
17	SB-M (44)		
18	WB-I (41)	WB-I (157)	
21	WB-I (37)		
25	WB-I (25)		
29			WAIS, SB-L (110)
36	WAIS (54)		
40		WAIS (156)	WAIS, SB-LM (48)
53	WAIS-R (41)	WAIS-R (118)	WAIS (53)
63			WAIS, WJ-R (48)
67			WAIS, WJ-R (33)
72	WAIS-R, WJ-R (31)		

Note. Available sample size for specific tests is contained in parentheses. SB-L, SB-M, SB-LM = Stanford–Binet Forms L, M, and LM; WB-I = Wechsler–Bellevue Intelligence Scale Form I; WAIS = Wechsler Adult Intelligence Scale; WAIS-R = Wechsler Adult Intelligence Scale–Revised; WJ-R = Woodcock–Johnson Psycho-Educational Battery–Revised.

ELSEVIER

Journal of Clinical Epidemiology 61 (2008) 1018–1027

Journal of Clinical Epidemiology

Item response theory facilitated cocalibrating cognitive tests and reduced bias in estimated rates of decline

Paul K. Crane[a,*], Kaavya Narasimhalu[a], Laura E. Gibbons[a], Dan M. Mungas[b], Sebastien Haneuse[c], Eric B. Larson[c], Lewis Kuller[d], Kathleen Hall[e], Gerald van Belle[f]

[a]Department of Medicine, University of Washington
[b]Department of Neurology, University of California at Davis
[c]Center for Health Studies, Group Health Cooperative
[d]Department of Epidemiology, University of Pittsburgh
[e]Department of Psychiatry, Indiana University – Purdue University in Indianapolis
[f]Departments of Biostatistics and Environmental and Occupational Health Sciences, University of Washington

Accepted 18 November 2007

Abstract

Objective: To cocalibrate the Mini-Mental State Examination, the Modified Mini-Mental State, the Cognitive Abilities Screening Instrument, and the Community Screening Instrument for Dementia using item response theory (IRT) to compare screening cut points used to identify cases of dementia from different studies, to compare measurement properties of the tests, and to explore the implications of these measurement properties on longitudinal studies of cognitive functioning over time.

Study Design and Setting: We used cross-sectional data from three large ($n > 1000$) community-based studies of cognitive functioning in the elderly. We used IRT to cocalibrate the scales and performed simulations of longitudinal studies.

Results: Screening cut points varied quite widely across studies. The four tests have curvilinear scaling and varied levels of measurement precision, with more measurement error at higher levels of cognitive functioning. In longitudinal simulations, IRT scores always performed better than standard scoring, whereas a strategy to account for varying measurement precision had mixed results.

Conclusion: Cocalibration allows direct comparison of cognitive functioning in studies using any of these four tests. Standard scoring appears to be a poor choice for analysis of longitudinal cognitive testing data. More research is needed into the implications of varying levels of measurement precision.

IDA Longitudinal: Next Steps

- Need for quantitative comparison across studies and better understanding of properties of particular measures
- International harmonization platform including metadata catalogue and harmonization rules (logical, psychometric)
  - Migrate IALSA search tool to the IHP catalogs (www.p3gobservatory.org)
  - Develop “Longitudinal Studies of Aging Data Schemas” that will facilitate harmonization across studies.
- Collect item-level data for psychometric harmonization
  - Identify item-level data required for co-calibration
  - Item library for quantitative comparison across past, current, and future studies: Map measures from existing studies with NIH Toolbox, HRS/SHARE, NIA-supported studies, international longitudinal studies

Co-Calibration and Development of "Item Library"

- Quantitative measurement harmonization
  - IRT and Latent Variable Approaches
  - Item library maps measures to one another
  - Requires measurement invariance across cohort, country, language
- Calibration requires common multivariate item sets across samples
  - Core Items
  - Planned missingness design can alleviate respondent burden
  - Use of bilingual subsamples permits cross-country harmonization; essential for direct comparison (e.g., WJ-R)

Strengths: Item Library Approach

- Retrospective Harmonization: Linkage across studies
  - Permits comparison of past, current, and future studies (i.e., cohort / social change)
    - Difficult to achieve measurement "standardization" in either national or international context
  - Evaluation of measurement equivalence and commonality/uniqueness of particular indicators
  - Retains breadth and innovation in study-specific measurement by permitting item/scale mapping to common constructs across studies
  - Provides basis for selection of "optimal" items/scales for current and future studies

Summary

- Co-calibration and Item Library would provide a way to obtain a predicted score for each person on a latent/idealized construct by linking through the association of the particular measure available in the study and the idealized measure
- Harmonization is currently restricted to particular subsets of studies with existing overlap in constructs of interest
- Broad harmonization efforts will require new data collection. If additional data were collected for co-calibration, there would be (at least) two options:
  - Collect data similar to what is available to support/improve harmonization across target studies
  - Collect data within ongoing studies where overlap is very thin/non-existent to maximize the number of additional studies that could be included.
- International harmonization efforts require sensitivity to language and form differences and efforts to establish measurement equivalence (e.g., using bilingual samples)