Methods for implementing CAR.

# Introduction

[Replication is ever more important, but rare and tricky for longitudinal designs]

Replicability of research findings in psychology recently came under increasing scrutiny (e.g. Maxwell et al, 2015, ADD). Multiple research findings failed to replicate (e.g. power pose, mindfulness meditation, etc, ADD). Calls for and attempts at systematic replication became more frequent (Open Science Collaboration, 2015). However, most of recent replication attempts involve **cross-sectional** designs and involve repeated data collection based on the original research protocols. Replications based on **longitudinal** designs (e.g. Brown et al, 2012; Piccinin et al, 2013, etc) are rare due to methodological and technical challenges and almost never involve collection of new data, instead relying on comparing analyses using existing longitudinal studies.

[Two broad classes of replication for longitudinal designs: pooling vs independent]

There are two broad classes of replications involving observational longitudinal data: those that pool data and those that do not. The goal of *pooling* approaches is to combined the data from different longitudinal studies and render them similar enough to analyze them as a single data set. Although powerful, this approach is too dependent on measurement harmonization, which is frequently unfeasible in area of cognitive aging. On the other hand, *non-pooling* approaches, coordinate analyses in independent studies to make results as comparable as possible.

[Degree and locus of control in coordinated analysis]

Traditional meta-analysis has the least control over its components. CAR expands this control, allowing iterative correction and improvement.

Hofer and Piccinin (2009) introduced the concept of *coordinated analysis with replication*, a research model for integrative data analysis.

# Portland workshop

In February of 2015 IALSA conducted a collaborative workshop in Portland, OR, bringing together the data from nine longitudinal studies[\*1]. Each of the study asked its data the same question: *is decline in physical abilities related to the decline in cognitive abilities as people age in later life*? The identical statistical operationalization of this question tested the data from each study: the bivariate growth curve model. While sharing no person-level data, each study submitted the results of model estimation (namely, Mplus output files) to the collective. The model pool, hosted in a public cloud location for transparency and accountability, counted over 1000 outputs, pairing 3 physical measures (pulmonary function, gait speed, and grip strength) with more than 50 different cognitive tests (e.g. mmse, block design, line orientation, etc.). This paper describes the coordinated analysis with replication (CAR) approach: a vehicle for organizing and implementing a large-scale replication of longitudinal analyses across multiple studies.

The CAR method aims to gather corroborating evidence for a broad hypothesis from multiple data sources. For example, the hypothesis "decline in pulmonary function is associated with decline in perceptual speed" may be tested by observing a statistical relationship between the trajectories of peak expiratory flow (pef) and performance on the symbol substitution task (symbol) in the data from Einstein Aging Study (EAS). The conclusion, however, even though supported  by statistical significance, would be limited to these specific measures and this specific sample. [ discuss limitation of pooled and meta analysis and transition to CAR’s response to the same challenges]

# Research Process

## 1. Research Proposal

Maelstrom and preliminary discussion.

## 2. Protocol Development

The preliminary planning settled on the bivariate growth curve model as the stencil for coordination.

## 3. Analysis Plan

We decided to conduct analysis individually for male and female subgroups because […] and in order to accommodate those studies, which collected data for a single gender (NAS). The set of covariates (measured at study entry) shared by all studies included age in years, years of education, height, smoking, diabetes, and cardiovascular disease. The last three measures were harmonized as a binary variable indicating the history at the time of study’s entry.

## 4. Statistical Analysis

Bivariate growth curve model was fitted to data from each longitudinal data. The model types include univariate (u0, u1, u2) and bivariate forms (b0,b1,b2), which estimated intercept only (u0, b0), intercept + linear slope (u1,b1) and intercept + linear slope + quadratic slope (u2, b2) models. Each of the form was controlled for six sets of covariates

## 5. Comparison of Results

Model outputs were upload to public github repository. We developed a suite of R script to parse and tabulate relevant indices.

## 6. Dissemination of Results

# Infrastructure

## Maelstrom

We used Maelstrom catalog during planning stage to find the optimal overlap among the studies in terms of available outcomes and covariates.

## Face-to-face

[Face-to-face as the forum to finalize data preparation]

Preliminary planning that involved utilization of Maelstrom online catalog, as well as extensive direct communication among the core team. However, many of the decisions regarding coordination of measurements and analyses required an extensive knowledge of the individual-level data and called for the focused attention of data managers. The face-to-face phase of Portland workshop spanned over 3 days and involved extensive discussion of studies’ idiosyncrasies of designs and measurements. Intermittent with discussion the final preparation of data took place: participants agreed upon the details of the operationalization for the harmonized variables and were able to check their implementation with one another.

[Coordinated analysis calls for common set of modeling skills]

Many of the researchers representing independent studies were early-career scientists, postdocs, or graduate students. The training component of the event allowed to level out the analytic skills of the group, ensuring that each participant was capable of implementing the estimation and basic interpretation of the fitted models.

[Benefits of face-to-face at early stage of the project]

While it was possible to convene the group via some virtual platform (Skype, WebEx, google, etc), we must advocate the utility of the face-to-face meeting. 1) Introduces people to each other 2) helps focus on the task at hand for a prolong period of time while limiting distractions 3) increases the feedback loop among participant during collective decision making 4) creates fellowship and establishes group self-identification by common goals 5) creates social momentum for follow up meetings

More personal, easier to get commitment –

## Follow-up

Some of the coordinating

## GitHub

Public github repository was used to 1) host submitted model outputs 2) coordinate follow-up activities 3) develop and host processioning and analytic scripts with Git version control

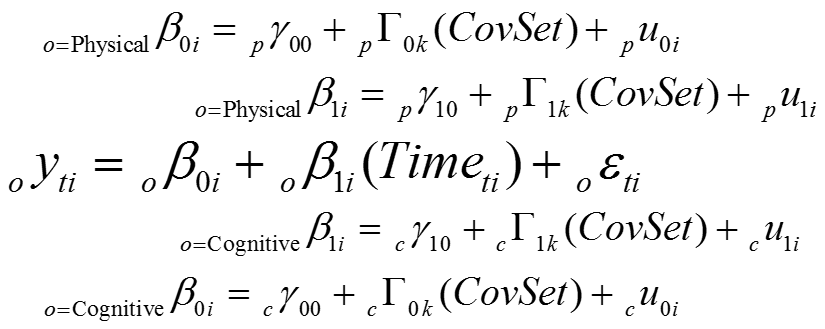
Advice for future events:

Andrea: better if we had one meeting before

* Have one meeting,
* Redcap – finish the survey before buying the ticket.

# Coordination

## Analysis

[model type] 

[subgroup]

Male, female. Reasons for subgroups.

[covariates]

Progressive sets

[measures]

## Measures

### Physical

[pef, fev]

[grip]

[gait, tug]

## Cognitive

[visual discrimination]

[perceptual speed]

[fluency]

[attention]

[fluid reasoning]

[mental status]

[executive function]

[working memory]

[short-term memory]

[semantic memory]

[episodic memory]

[verbal comprehension]

# Ingredients

* Which acronym is better/more telling : CAR (vehicle for progress, driver) or CAWR (core)?
* DataSHIELD apprears to bet on the technological solution, while CAWR makes much heavier use of personal collaboration.
* DataSHILD is a high-tech solution, while CAWR is a low-tech solution

# Unfit paragraphs

[There are many proposed operationalizations to replication for longitudinal designs.]

A number of approaches for replicating findings from longitudinal studies have been conceptualized. *Sequential independent replication* (e.g. Vand Dijk et al, 2008) examines published findings and evaluates other independent longitudinal studies for supporting evidence to the original claim. *Meta-analysis* (e.g. Sutton & Higgins, 2008) combines standardized effects from a set of published findings and estimates the general effect. Another class of approaches combines individual-level data from multiple studies (e.g. data pooling, integrative data analysis, mega-analysis, pooled data meta-analysis, etc). One needs to make a lot of decisions to operationalize what “replication” means in longitudinal setting, hence the variety of proposed solutions.

# References

[1] Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does “failure to replicate” really mean?. *American Psychologist*, *70*(6), 487.

[2] Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716.

Roediger, H. L. (2012). [Psychology’s woes and a partial cure: The value of replication](http://www.psychologicalscience.org/observer/psychologys-woes-and-a-partial-cure-the-value-of-replication#.WHPDvfkrJhF). *APS Observer*, *25*(2), 9.

Piccinin, A. M., Muniz-Terrera, G., Clouston, S., Reynolds, C. A., Thorvaldsson, V., Deary, I. J., ... & Starr, J. M. (2013). [Coordinated analysis of age, sex, and education effects on change in MMSE scores](http://psychsocgerontology.oxfordjournals.org/content/68/3/374.short). *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *68*(3), 374-390.

Brown, C. L., Gibbons, L. E., Kennison, R. F., Robitaille, A., Lindwall, M., Mitchell, M. B., ... & MacDonald, S. W. (2012). [Social activity and cognitive functioning over time: a coordinated analysis of four longitudinal studies](http://downloads.hindawi.com/journals/jar/2012/287438.pdf). *Journal of aging research*, *2012*.