# Portland CAR: next steps

2016-08-05

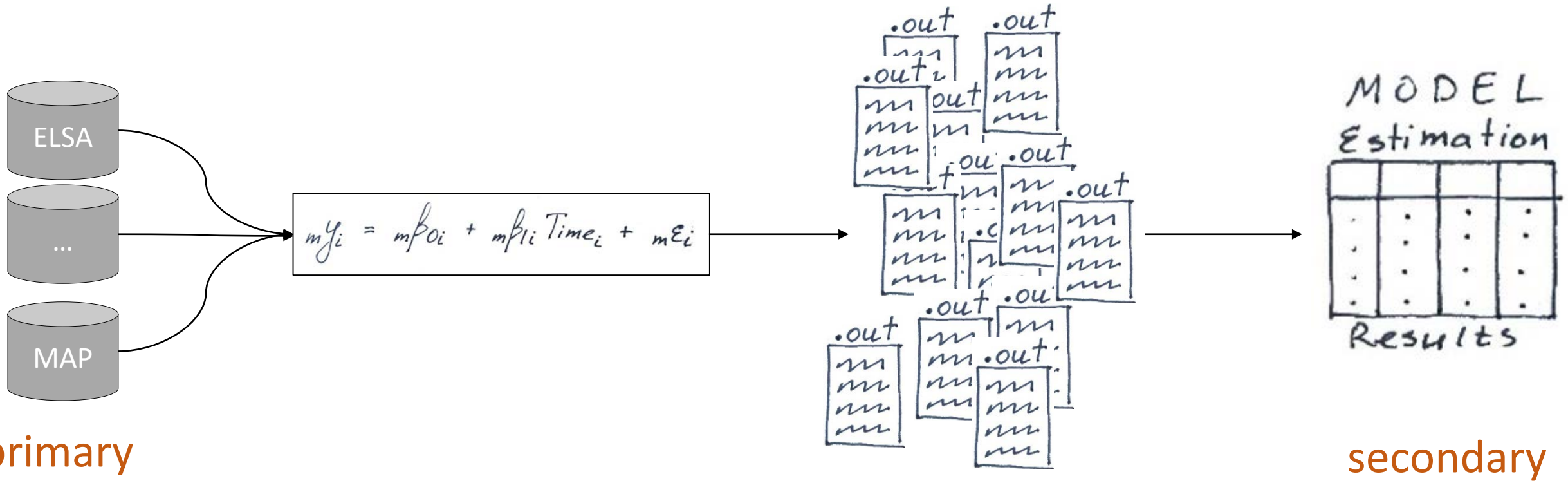
# Contents

- Preamble : context so far
- CPA poster : Gen2 scripts
- Next steps

# Original vision

Aggregation Phase: Coordinate statistical analysis of the primary data

Analysis Phase: Facilitate the analysis of the secondary data

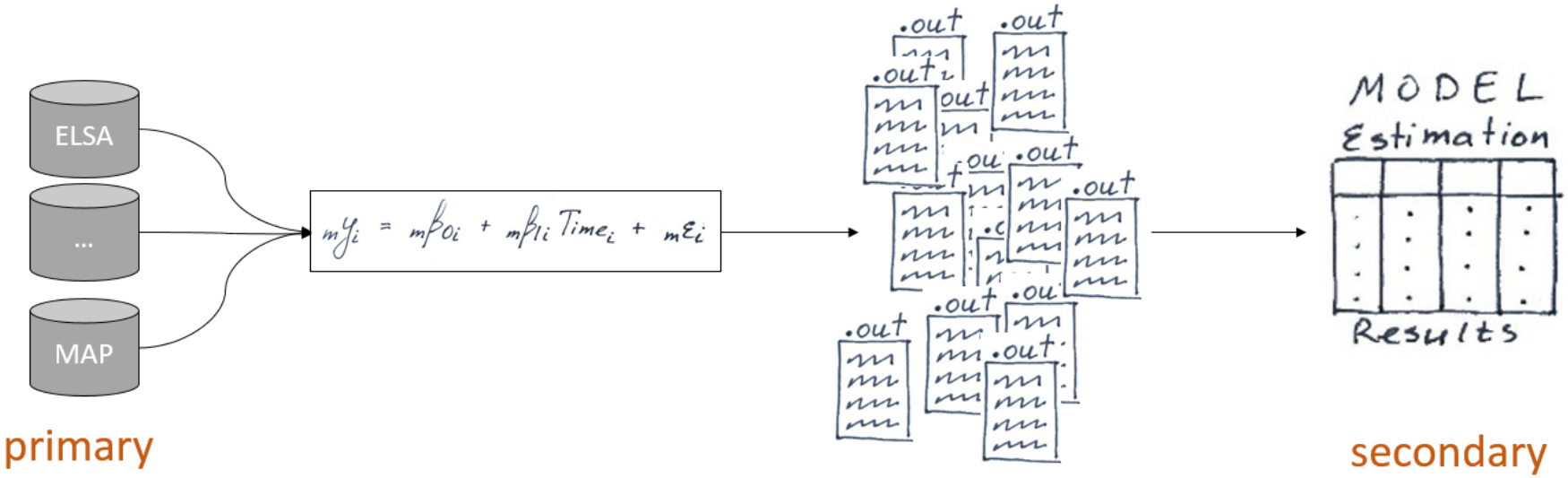



primary

secondary

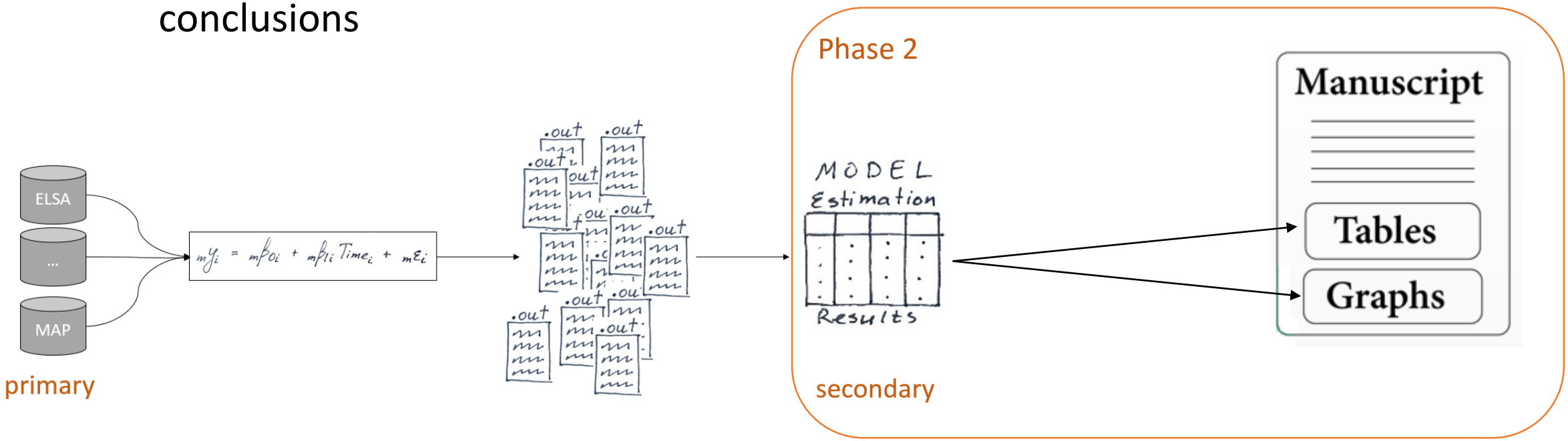
# Original vision

1. Coordinate statistical analysis of the primary data
   A. Standardize input data
   B. Standardize syntax for fitting models
   C. Remove human error during extracting indices from outputs
2. Facilitate the analysis of the secondary data



# Original vision

1. Coordinate statistical analysis of the primary data
2. Facilitate the analysis of the secondary data
   - Originally we thought a simple BISR correlation will be all we need
   - But there were many unexpected specifics that threatened the validity of conclusions



# Current progress in the Analysis Phase

- Table of BISR correlation
- Table of Growth processes
- Study-specific tables for manuscript seeds

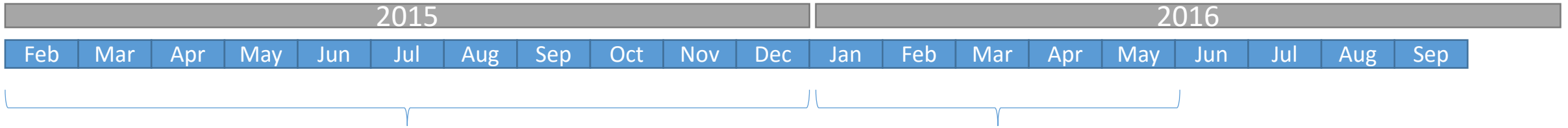
target process a : gait

| | Fixed effects | animals | bnt | mmse | gait |
|---|---|---|---|---|---|
| a | intercept | 18.93(2.41),** | 18.93(2.41),** | 18.93(2.41),** | 18.93 |
| a | slope | 0.80(0.47)* | 0.80(0.47)* | 0.80(0.47)* | 0.08 |
| a | age | -0.15(0.16)** | -0.15(0.16)** | -0.15(0.16)** | -0.15 |
| a | education | 0.88(0.26)*** | 0.88(0.26)*** | 0.88(0.26)*** | 0.88 |
| a | height | -0.01(0.11) | -0.01(0.11) | -0.01(0.11) | 0 |
| a | smoking | 1.53(1.44) | 1.53(1.44) | 1.53(1.44) | 1.15 |
| a | cardio | -0.38(2.44)* | -0.38(2.44)* | -0.38(2.44)* | 0 |
| a | diabetes | -4.39(2.59)*** | -4.39(2.59)*** | -4.39(2.59)*** | -4.39 |
| a | slope*age | -0.04(0.03)** | -0.04(0.03)** | -0.04(0.03)** | -0.04 |
| a | slope*education | -0.07(0.05) | -0.07(0.05) | -0.07(0.05) | 0 |
| a | slope*height | -0.01(0.02) | -0.01(0.02) | -0.01(0.02) | 0 |
| a | slope*smoking | -0.10(0.26)* | -0.10(0.26)* | -0.10(0.26)* | 0 |
| a | slope*cardio | -0.07(0.73)*** | -0.07(0.73)*** | -0.07(0.73)*** | -0.07 |
| a | slope*diabetes | 0.18(0.40) | 0.18(0.40) | 0.18(0.40) | 0.01 |
| b | intercept | 38.58(2.78)* | 38.58(2.78)* | 38.58(2.78)* | |
| b | slope | -0.78(0.72)*** | -0.78(0.72)*** | -0.78(0.72)*** | |
| b | age | -0.35(0.19)*** | -0.35(0.19)*** | -0.35(0.19)*** | |
| b | education | 0.72(0.29) | 0.72(0.29) | 0.72(0.29) | |
| b | height | -0.10(0.14) | -0.10(0.14) | -0.10(0.14) | |
| b | smoking | 0.78(1.34) | 0.78(1.34) | 0.78(1.34) | |
| b | cardio | 1.06(3.11) | 1.06(3.11) | 1.06(3.11) | |
| b | diabetes | -5.08(2.30)**' | -5.08(2.30)**' | -5.08(2.30)**' | |
| b | slope*age | -0.03(0.04)** | -0.03(0.04)** | -0.03(0.04)** | |
| b | slope*education | 0.07(0.06) | 0.07(0.06) | 0.07(0.06) | |
| b | slope*height | 0.00(0.03) | 0.00(0.03) | 0.00(0.03) | |
| b | slope*smoking | 0.08(0.39)* | 0.08(0.39)* | 0.08(0.39)* | |
| b | slope*cardio | 0.00(0.76) | 0.00(0.76) | 0.00(0.76) | |
| b | slope*diabetes | 0.01(0.66) | 0.01(0.66) | 0.01(0.66) | |
| | Variance components | | | | |
| a | Intercept | 6.13(.52)** | 6.13(.52)** | 6.13(.52)** | |
| a | Slope | .85(.12)* | .85(.12)* | .85(.12)* | |
| a | Residual | 1.23(1.02)*** | 1.23(1.02)*** | 1.23(1.02)*** | |
| b | Intercept | 3.34(.52)*** | 3.34(.52)*** | 3.34(.52)*** | |
| b | Slope | .66(.08)* | .66(.08)* | .66(.08)* | |
| b | Residual | 2.08(.82)* | 2.08(.82)* | 2.08(.82)* | |
| | Covariance Components | | | | |
| | Cov(IPhys-ICog) | 0.076(.051) | 0.076(.051) | 0.076(.051) | |
| | Cov(SPhys-SCog) | 0.076(.051) | 0.076(.051) | 0.076(.051) | |
| | Cov(IPhys-SCog) | 0.076(.051) | 0.076(.051) | 0.076(.051) | |
| | Cov(ICog-SPhys) | 0.076(.051) | 0.076(.051) | 0.076(.051) | |
| | AIC | 44259.46 | 44259.46 | 44259.46 | |
| | BIC | 44345.68 | 44345.68 | 44345.68 | |

This presentation is about phase of
mechanical aggregation.

The subjective analysis phase is for another meeting

# Two generation of software scripts



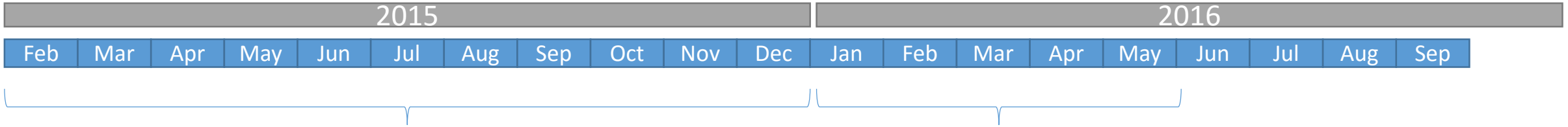

Generation 1

Generation 2

Limitations

- Difficult to re-run models
- Humans create syntax files
- No certainty that data described are data modeled

# Two generation of software scripts



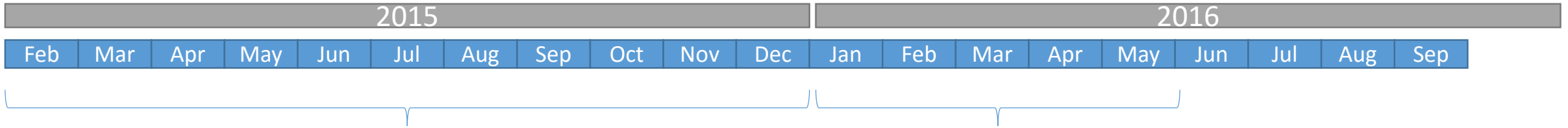

Generation 1

Generation 2

Uncertainty about:

- Right subjects? (e.g. wrong subgroup filter)
- Misspecified models? ( e.g. relied on filename for model shape)
- Violated convention that were suggested to the drivers?

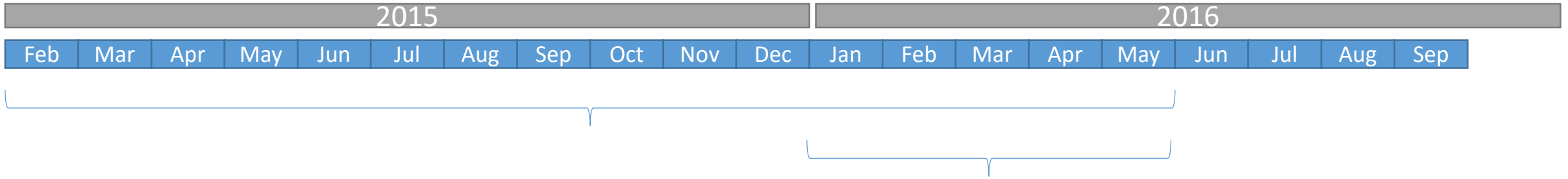
# Two generation of software scripts




Creates Threats to Statistical Conclusion Validity (Analysis Phase):

1. Many models have not estimated correlations
   - (instead, computed post hoc)
2. We cannot rely ONLY on the value of BISR correlations. Here's why:
   A. Variance of slope
   B. Sample size due to subgroup split
      A. What if we drop models with insufficient sample size? (e.g. N < 100 )
   C. Number of included waves
   D. Untraced human errors during estimation

# Two generation of software scripts




Generation 1

Limitations

- Difficult to re-run models
- Humans create syntax files
- No certainty that data described are data modeled

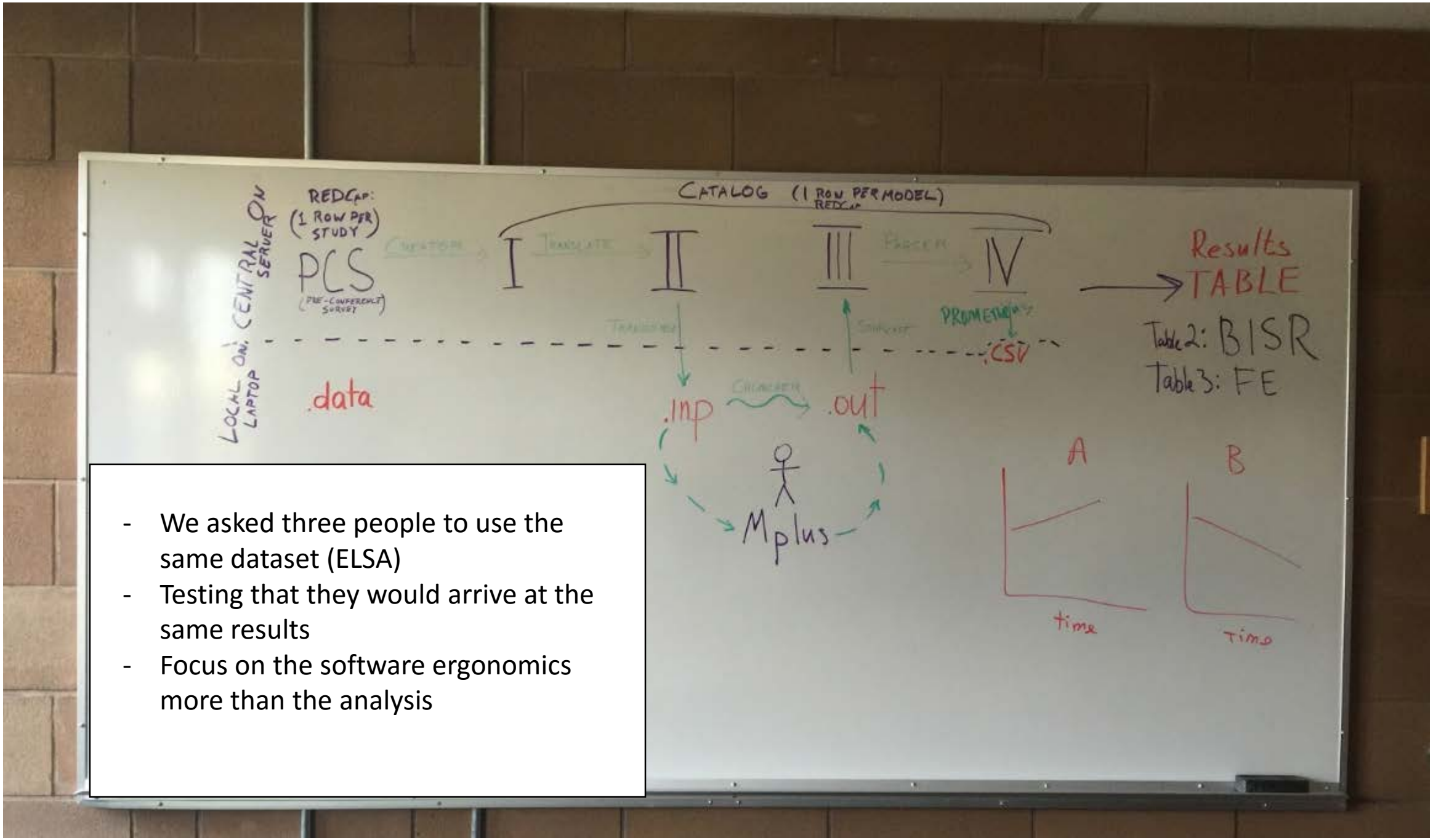
Generation 2

Must improve

- Easy to re-run models
- Automatic generation of syntax files for each study
- Ensure the data described is the data modeled
- Greater trust in replication

On May 26, 2016 we have successfully tested aggregation phase or Gen 2 scripts using ELSA data

UVic May 26, 2016




- We asked three people to use the same dataset (ELSA)
- Testing that they would arrive at the same results
- Focus on the software ergonomics more than the analysis

# POSTER PRESENTATION

# (Automated) Chain of Custody

- evidence must be documented, otherwise it can’t be used in courtroom
- cannot be vouched for
- can be contaminated during investigation

Gen 2 has the evidence under control the entire time from the crime scene to the courtroom

- Adds transparency and reproducibility to the process
- Videotaping the entirety of the investigation
- No assurance that the knife found on the crime scene is the murder weapon.
- But solid confidence that the knife presented in the courtroom is the knife found at the crime scene

# (Automated) Chain of Custody

Start:

- Pre-Conference Survey + data
- No human intervention after that:
  - no subjective decision
  - only click-and-run
  - oversight of script execution

Without ACC we cannot be certain about:

- Right subjects? (e.g. apply subgroup filter)
- Misspecified models? (relied on filename for model shape)
- Violated convention that were suggested to the drivers?

# Bottlenecks

## 1 generation

- generating and aggregating the output ( ~12 months)
- comprehending the aggregated output ( ~ 5 months into it)

## 2 generation

- ~~generating and aggregating the output (~ 1 day)~~
- comprehending the aggregated output

# Advantages of Gen 2

- Lower cost of collaboration during coordinated analysis
- Alleviates the disheartening difficulty/length of result extraction
- If each workshop takes too long to process then you will be tempted to swing for the fences (become more aggressive to achive results here and now, b/c you don’t want to wait for another 18 months.
- greater focus on achievable goals

# Advantages of Gen 2

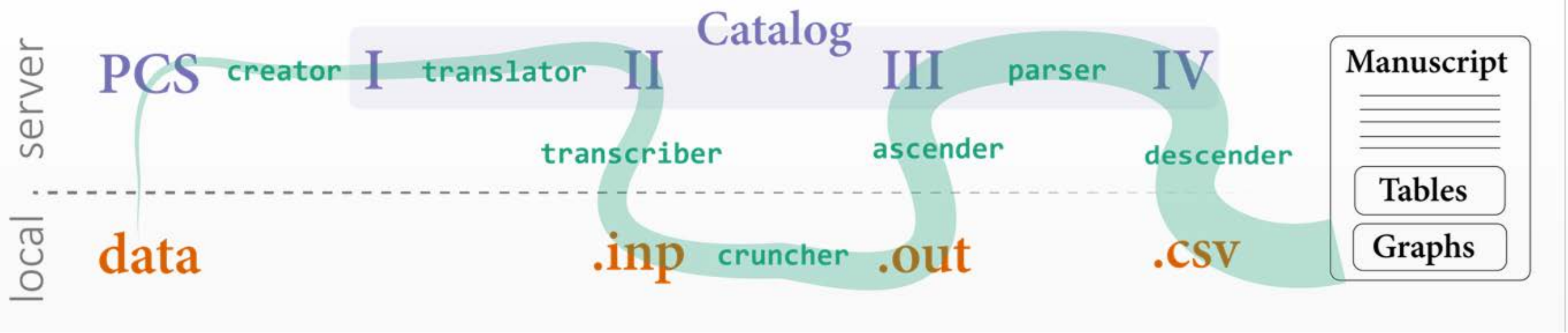
Focus changes

- From :How many models can we bring together?
- To:  How can we organize the results?
- To : How many results can we make sense of?

- Will take less time
- Can do remotely
- More frequent
- More focused workshop
- Greater emphasis on Phase 2

# Future directions

1. iLifeSpan-based:
   - Groom available studies to fit Portland needs = standard for a “general grooming”
2. Same model (BISR), new workshop with new studies or/and variables
   1. Keep variables, change studies
   2. change variables, change studies
3. New statistical model
   1. Old studies
   2. New studies


Catalog
server
local
PCS
creator
I
translator
II
III
parser
IV
transcriber
ascender
descender
data
.inp
cruncher
.out
.csv
Manuscript
Tables
Graphs

# IALSA-study-curator project

| Study | | |
|---|---|---|
| Einstein Aging Study | EAS | |
| English Longitudinal Study of Aging | ELSA | 100% |
| Health and Retirement Study | HRS | |
| Interdisciplinary Longitudinal Study of Aging | ILSE | |
| Normative Aging Study | NAS | |
| Quebec Longitudinal Study on Nutrition and Aging | NuAge | |
| Octogenarian Twins | OCTO | |
| Rush Memory and Aging Project | MAP | 100% |
| Swedish Adoption Twin Study of Aging | SATSA | |
| Longitudinal Aging Study Amsterdam | LASA | |
| Whitehall II | WH | |

# New options in Gen 2

- Unlike Gen 1 that offered only option 1, Gen 2 offers different types of workshops:

- together ( 4 days + travel costs)
- remote completely (regular, spaced out meetings online)
- hybrid (muscle happens quickly, more time dedicated to the interpretation and writing), happens at the conference workshop ( ~6 hours