

Writing Simulators: “There’s more to it than the stats analysis” – advice to statisticians from a computer programmer



Tom Parke
Presented at PSI UK
April 27th

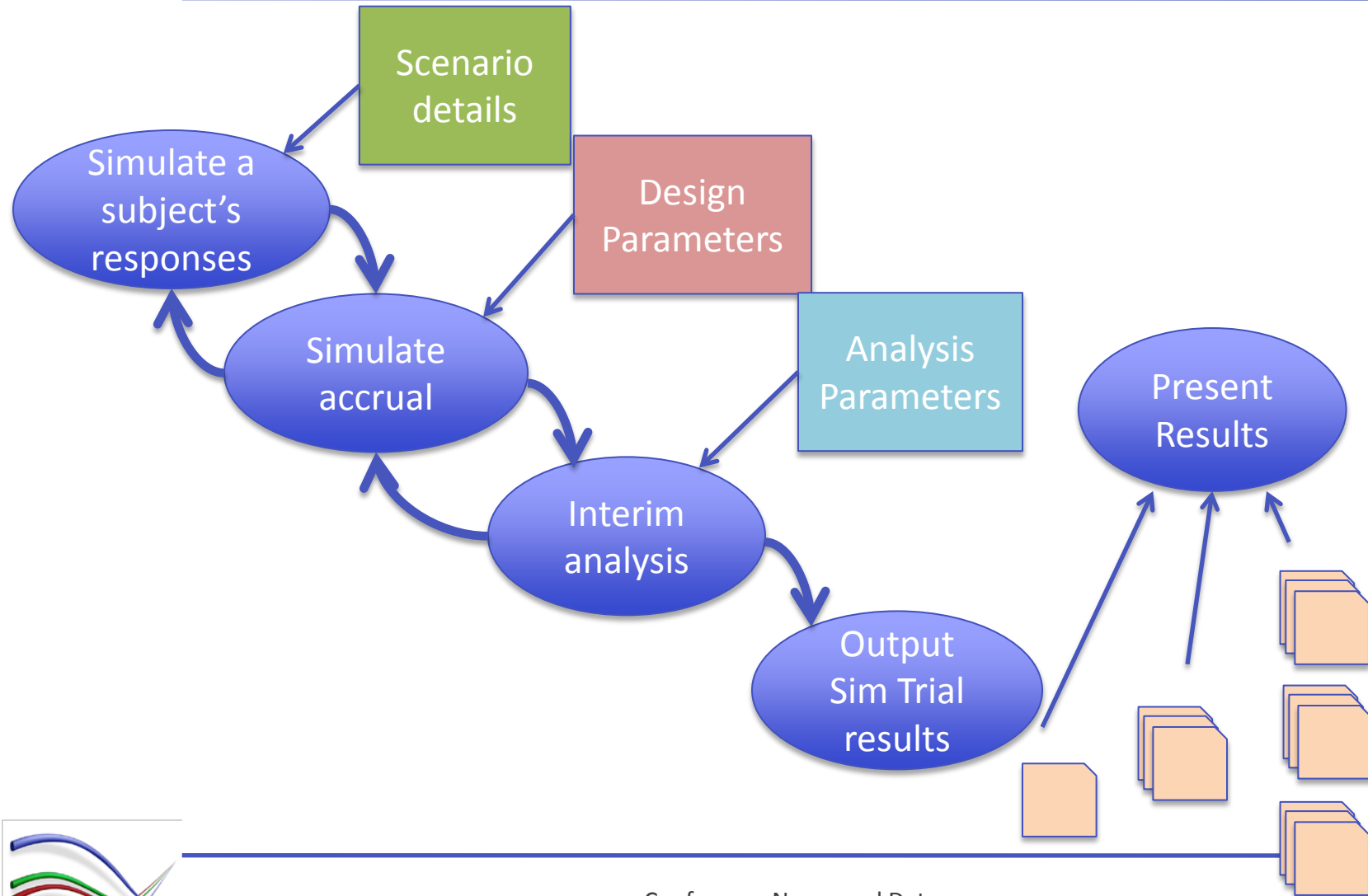
Writing a Clinical Trial Simulator

- Background:
 - working on trial simulators since 2002: Pfizer, MD Anderson, Wyeth, Eli Lilly ... FACTStm
 - Worked on R simulations for ISCTM & DIA Adaptive Program Working Group
 - Before that worked in SW since 1979 on compilers, operating systems, real time control s/w, air traffic control s/w, internet domain name registry and an options trading system.

The Problem

- The simulation results need to be trusted
- The results need to be available when required
- The simulation will need to change as the plan changes
- You will need to back months later and remember where results came from (reproducibility)

Analysis -> Simulation



Analysis -> Simulation

- The whole thing has got bigger
- But the stuff round the analysis is just “book keeping”
- Yes, but the chance of making an error in software is not strongly correlated to how intellectually challenging that s/w is.
- If the stuff around the analysis (which you will not find so interesting) is not given the time and attention it requires you will fail

Failure ...

- The results will not be trusted
- The results will not be available on time
- The simulation will not be able to be easily adapted as the plan changes
- Months later it will not be clear exactly how simulation results were generated

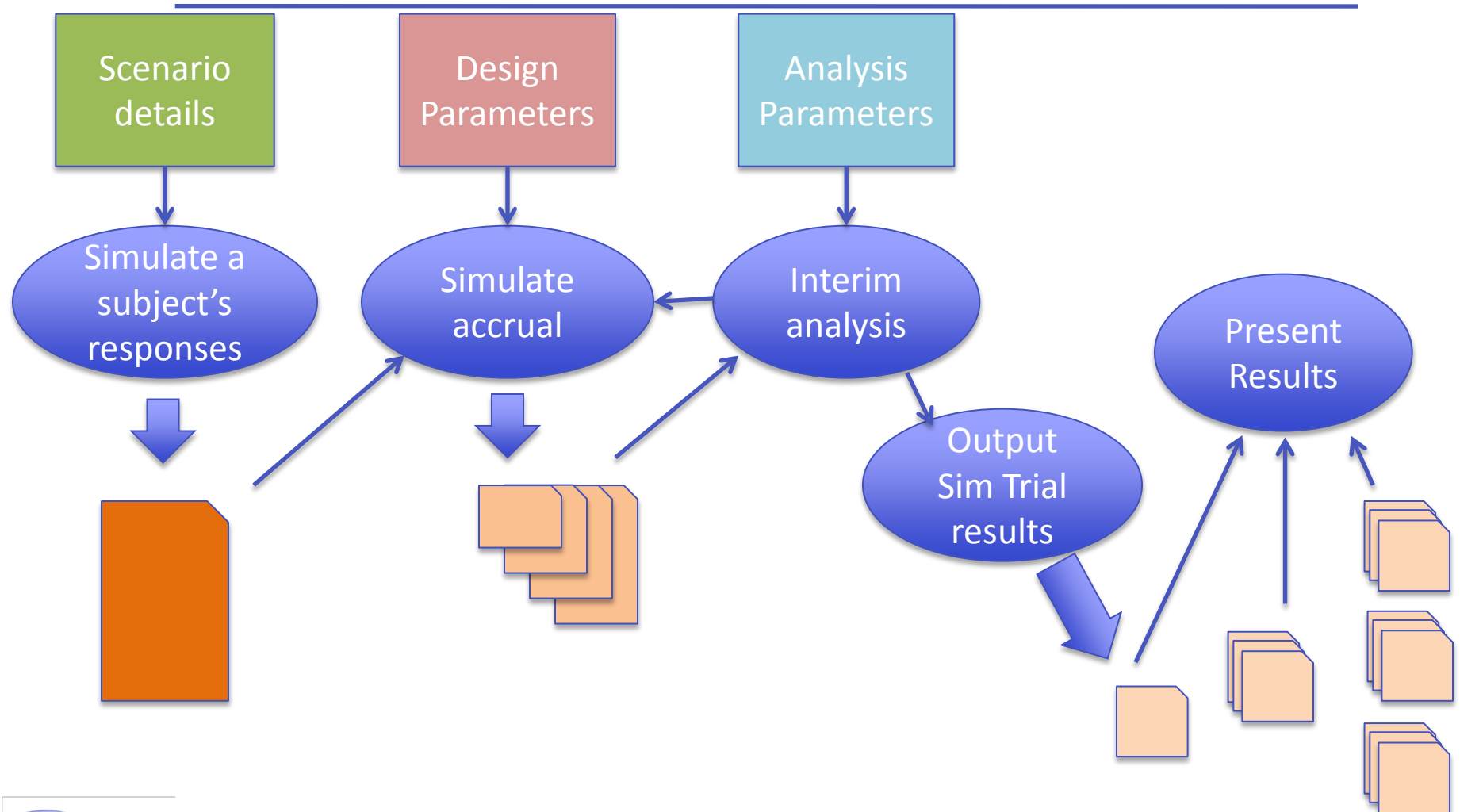
Truisms

- The simulations will be run and re-run far more often than you expected
- The simulations will have to be modified far more often than you expected
- Successful software is run for a lot longer than you expected
- But most software is not successful.....

How to tackle this

- Modular approach
- Start Simple
- Test as you go
- Retain intermediate results
- Attend to simulation hygiene

Modular Approach



Benefits of modularity

- Can build and test a stage at a time
- Can run a stage at a time and re-run a stage
- When we have to make changes we will often be able to isolate them to a particular stage
- Start with a simple analysis stage and get everything else working first.

Start Simple

- Postpone the intellectually challenging (a.k.a interesting) part of the project to later
- Develop the “book keeping” code first
- Give the BK code the attention it requires
- Get end-to-end results with simple analysis
- Incrementally build up the interesting part of the project - compare results with each increment.

Testing

- Test components individually
- Test Single analyses first (not averages over many sims)
- Be especially suspicious of early “good” results!
- Chase down any discrepancy, no matter how minor

Hygiene

arm 300.facts	arm 330_results	arm 330.facts	arm 360_results	arm 360.facts	arm 390_results	arm 390.facts	arm ni 150_results	arm ni 150.facts	arm ni 180_results	arm ni 180.facts	arm ni 210_results
Phase 2 fixed 3-arm ni 210.facts	Phase 2 fixed 3-arm ni 240_results	Phase 2 fixed 3-arm ni 240.facts	Phase 2 fixed 3-arm ni 270_results	Phase 2 fixed 3-arm ni 270.facts	Phase 2 fixed 3-arm ni 300_results	Phase 2 fixed 3-arm ni 300.facts	Phase 2 fixed 3-arm ni 330_results	Phase 2 fixed 3-arm ni 330.facts	Phase 2 fixed 80_results	Phase 2 fixed 80.facts	Phase 2 fixed 100_results
Phase 2 fixed 100.facts	Phase 2 fixed 120_results	Phase 2 fixed 120.facts	Phase 2 fixed 130_results	Phase 2 fixed 130.facts	Phase 2 fixed 140_results	Phase 2 fixed 140.facts	Phase 2 fixed 150_results	Phase 2 fixed 150.facts	Phase 2 fixed 160_results	Phase 2 fixed 160.facts	Phase 2 fixed 170_results
Phase 2 fixed 170.facts	Phase 2 fixed 180_results	Phase 2 fixed 180.facts	Phase 2 fixed 190_results	Phase 2 fixed 190.facts	Phase 2 fixed 200_results	Phase 2 fixed 200-smallsim_results	Phase 2 fixed 200-smallsim.facts	Phase 2 fixed 200.facts	Phase 2 fixed 220_results	Phase 2 fixed 220.facts	Phase 2 fixed 240_results
Phase 2 fixed 240.facts	Phase 2 fixed 260_results	Phase 2 fixed 280.facts	Phase 2 fixed 320_results	Phase 2 fixed 320.facts	Phase 2 multi-eq-int 3-ar...33_results	Phase 2 multi-int 3-arm 300 F33_results	phase 2 multi-int 100 3-ar...0_results	phase 2 multi-int 100 3-arm 240.facts	phase 2 multi-int 100 3-ar...0_results	phase 2 multi-int 100 3-arm 270.facts	phase 2 multi-int 100 3-ar...0_results
phase 2 multi-int 100 3-arm 300.facts	phase 2 multi-int 120 3-ar...0_results	phase 2 multi-int 120 3-arm 240.facts	phase 2 multi-int 120 3-ar...0_results	phase 2 multi-int 120 3-arm 270.facts	phase 2 multi-int 120 3-ar...0_results	phase 2 multi-int 120 3-arm 300.facts	phase 2 multi-int 140 3-ar...0_results	phase 2 multi-int 140 3-arm 240.facts	phase 2 multi-int 140 3-ar...0_results	phase 2 multi-int 140 3-arm 270.facts	phase 2 multi-int 140 3-ar...0_results
phase 2 multi-int 140 3-arm 300.facts	Phase 2-3 simulati...lans.docx	sfn.3adapt.300.100.cdf	sfn.3ami120.270	Sims-delete-bug1.facts	Sims-delete-bug2.facts	sims.140.2int.csv	sims.160.5int.csv	standards.R	subroutines.f	tmp.csv	TTE-program-analysis-v14.R
phase 2 multi-int 140 3-arm 300.facts	Phase 2-3 simulati...lans.docx	sfn.3adapt.300.100.cdf	sfn.3ami120.270	Sims-delete-bug1.facts	Sims-delete-bug2.facts	sims.140.2int.csv	sims.160.5int.csv	standards.R	subroutines.f	tmp.csv	TTE-program-analysis-v14.R
v14 phase 2 adaptive...eval.csv	v14 phase 2 adaptive...eval.csv	v14 phase 2 adaptive...eval.csv	v14 phase 2 adaptive...eval.csv	v14 phase 2 adaptive...val A.csv	v14 phase 2 adaptive...eval.csv	V14-2arm-1int-080-details.csv	V14-2arm-1int-080.csv	V14-2arm-1int-100-details.csv	V14-2arm-1int-100.csv	V14-2arm-1int-120-details.csv	V14-2arm-1int-120.csv
v14 phase 2 adaptive...eval.csv	v14 phase 2 adaptive...eval.csv	v14 phase 2 adaptive...eval.csv	v14 phase 2 adaptive...eval.csv	v14 phase 2 adaptive...val A.csv	v14 phase 2 adaptive...eval.csv	V14-2arm-1int-080-details.csv	V14-2arm-1int-080.csv	V14-2arm-1int-100-details.csv	V14-2arm-1int-100.csv	V14-2arm-1int-120-details.csv	V14-2arm-1int-120.csv
V14-2arm-1int-140-details.csv	V14-2arm-1int-140.csv	V14-2arm-1int-160-details.csv	V14-2arm-1int-160.csv	V14-2arm-1int-180-details.csv	V14-2arm-1int-180-details.xlsx	V14-2arm-1int-180.csv	V14-2arm-1int-180 sens-dtis-00.csv	V14-2arm-1int-180 sens-res.csv	V14-2arm-1int-200-details.csv	V14-2arm-1int-200.csv	V14-2arm-1int-240-details.csv
V14-2arm-1int-140-details.csv	V14-2arm-1int-140.csv	V14-2arm-1int-160-details.csv	V14-2arm-1int-160.csv	V14-2arm-1int-180-details.csv	V14-2arm-1int-180-details.xlsx	V14-2arm-1int-180.csv	V14-2arm-1int-180 sens-dtis-00.csv	V14-2arm-1int-180 sens-res.csv	V14-2arm-1int-200-details.csv	V14-2arm-1int-200.csv	V14-2arm-1int-240-details.csv

CONFIDENTIAL NAME AND DATE

Refactoring

- As you make changes
- Improve the code
- Not so much with complexity, but keeping it simple and clear

Performance

- For type-1 error you may need many sims
- But for design comparison & optimization you need fewer – we want ‘significant’ improvement due to large difference not small SE!
- But incredibly powerful to be able to run ‘indicative’ number of sims (100-1,000) during a meeting, i.e. in a few minutes.

If using R

- Learn how to avoid loops
 - Vector operations
 - Apply() family of functions

Do as I say ... ;-)

- It is easy to not start out on the one true way
- And easy to be drawn away from it
- But it always, always pays to try to get back on it.