JENNIFER THOMPSON

American Hospital Association
October 5, 2018

(PROFESSIONAL) ME: A QUICK INTRODUCTION

EDUCATION

MPH in Biostatistics

Workshops, short courses, trainings in statistics & programming

EXPERIENCE

13+ years using data, analysis, communication to drive science forward, improve patient care + experience

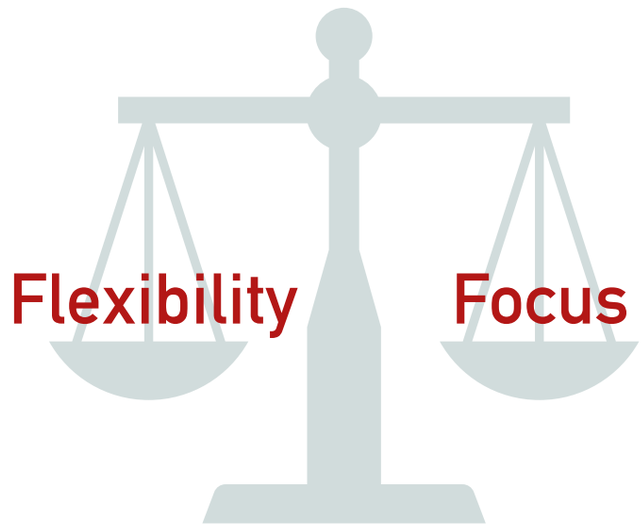
Collaboration
Data wrangling & rigorous analysis
Reproducible, transparent practices
Clear, concise communication
Project management
Mentoring, education

ENTHUSIASM

MY APPROACH TO RESEARCH

MY APPROACH TO RESEARCH

1. Define the relevant question(s)
   - ➤ Which answers are **actionable** - would change practice, increase understanding?
   - ➤ Which questions do we have **resources** to reliably answer?
2. Form a plan
   - ➤ Convert questions to **measurable** outcomes
   - ➤ Develop study **design**, statistical **analysis plan**: as complex as needed, but no more
   - ➤ Include ongoing maintenance (eg, web app), dissemination of results
   - ➤ Use plan to manage ongoing project scope



MY APPROACH TO RESEARCH

3. Assess and connect **resources**

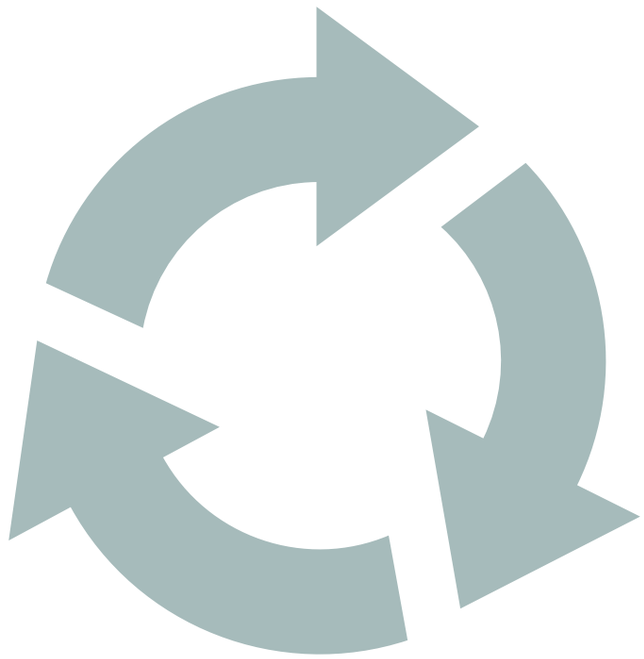
- Data
- Domain experts, other team members
- Analytic/presentation tools

4. **Wrangle & explore** data

- Are **assumptions** accurate?
- How should data be **transformed** to answer our questions?

5. **Execute** the plan

- Sounds deceptively simple!
- Communicate roadblocks to team, stakeholders
- Adapt (and document) as needed



MY APPROACH TO RESEARCH

6. Alongside entire team, interpret & communicate impact, results, methods

   - ➤ Need clear message without loss of nuance
   - ➤ Each level of detail important for different audience
   - ➤ Emphasize conciseness, visuals in product(s) appropriate for goals, audience
     (manuscript/white paper, web site/app, presentation…)

THROUGHOUT THE PROCESS

- Priorities reflect our goals & mission
- Data is complete, accurate
- Research team *understands* raw data
- Thorough, appropriate analysis
- Primary message & methods communicated appropriately for specific audience(s), whether by me or colleagues
  Can my clinical teammates explain our results to other clinicians?
  Can I explain real-world impact to other statisticians?
- Work done in reproducible, rigorous manner
  - Sensible organization and documentation
  - Literate programming/reproducible reporting
  - Version control
  - Aspirational: Code review

SCENARIO 1:

NIH R01 CLINICAL TRIAL

SITUATION 1: PROSPECTIVE, NATIONWIDE DATA COLLECTION

- ~15 sites, 6 years of enrollment -> staff turnover, fuzzy memories
- Complex, longitudinal study design -> detailed data collected by hand, with different data required in different study phases

SOLUTION

R script to find, document potential data issues in "real" time using database API

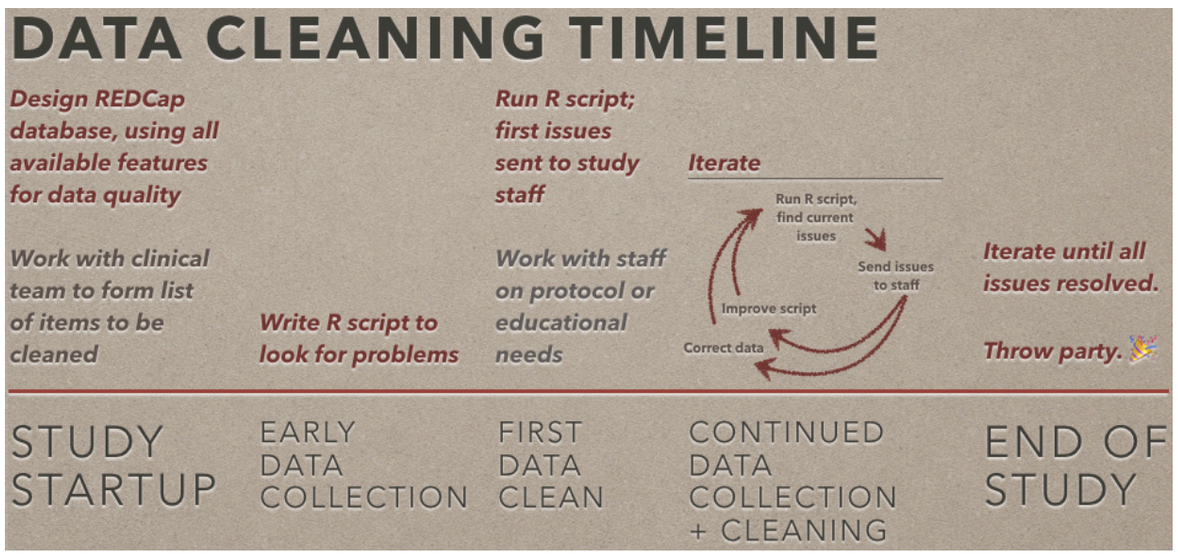
RESULTS

Me:
Efficient data cleaning
After initial investment, minor time involved

Clinical team:
Not "fun" but better!
Changes to original data well documented; less painful than leaving to study end

Data:
Better, faster
More accurate for interim analyses, DSMB
Ready for final analysis more quickly

Have trained others on this process; tutorial available on GitHub



SITUATION 2: REPRODUCIBLE, OPEN RESEARCH TAKING OFF

- Reproducibility is of growing importance in scientific community; open science becoming more and more the norm 👍
- Group doesn't have direct experience with tools and practices

SOLUTION

Spearhead efforts toward transparency, reproducibility for major publication

Training on Open Science Framework, become group lead

Ensure R code is organized, version controlled, on GitHub

Analysis in RMarkdown, interweaving code + results

RESULTS

Group **leads** area in adoption; results are more **trustworthy**

SITUATION 3: BALANCING COMPLEXITY AND CLARITY

- Primary outcome is a) complex, combining three related metrics, and b) has a notoriously bimodal distribution
- Best choice for modeling has a clinically difficult interpretation, especially combined with outcome's inherent complexity

Lead efforts to incorporate colleagues' topical statistical research into clinical work

SOLUTION

Clearer interpretation that maintains statistical **rigor**

RESULT

"adjusted odds (CI) of more vs fewer days alive & free of delirium and coma for [treatment] vs placebo"

"adjusted median (CI) days alive with normal brain function for each group"

SITUATION 4: COMPLEX, LONGITUDINAL STUDY DESIGN

Study design (& therefore data collection) is unusual & complex:

- Four study phases; patients could be disqualified within 5 days or randomized at varying times after consent
- Outcomes are summaries of raw data

SOLUTION

Rigorous data wrangling process

- Formal tests for raw & transformed data
- Detailed file structure + version control
- Thorough commenting, documen-tation

→ Reliable analysis datasets

→ Easier to identify, fix bugs

→ Easier to collaborate

RESULT

More **reliable**, transportable code + data

SCENARIO 2:

NIH COHORT STUDY

SITUATION: NEED TO MONITOR ENROLLMENT, STUDY CONDUCT

- Team, PI need regular updates on enrollment targets, protocol points, follow-up rates
- Need to serve variety of folks with broad range of backgrounds
- Previous: Detailed PDF; worked, but buried important metrics

SOLUTION

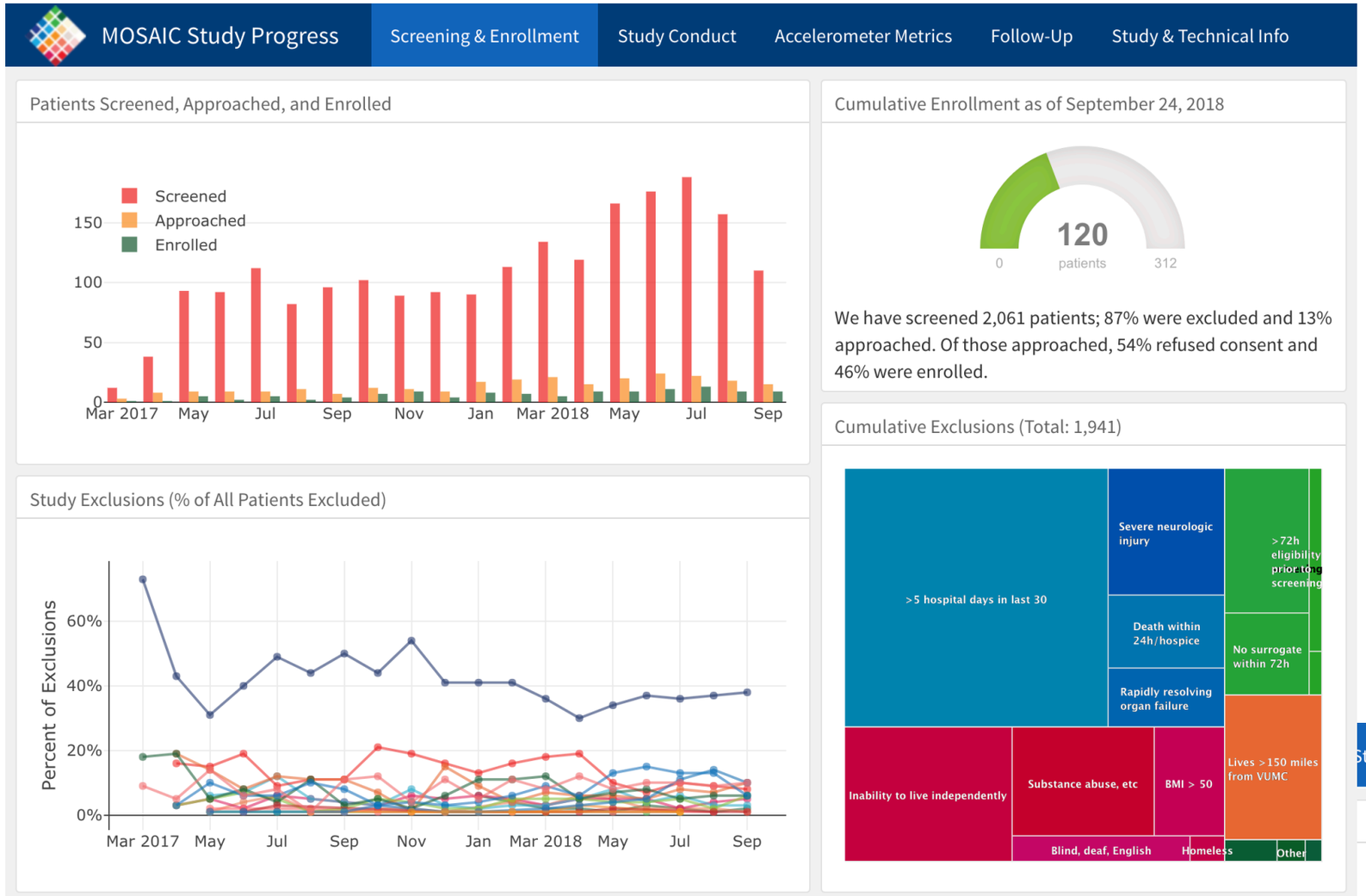
Dashboards for quick info, warning system for pain points

RESULTS

Quickly see successes, attention areas

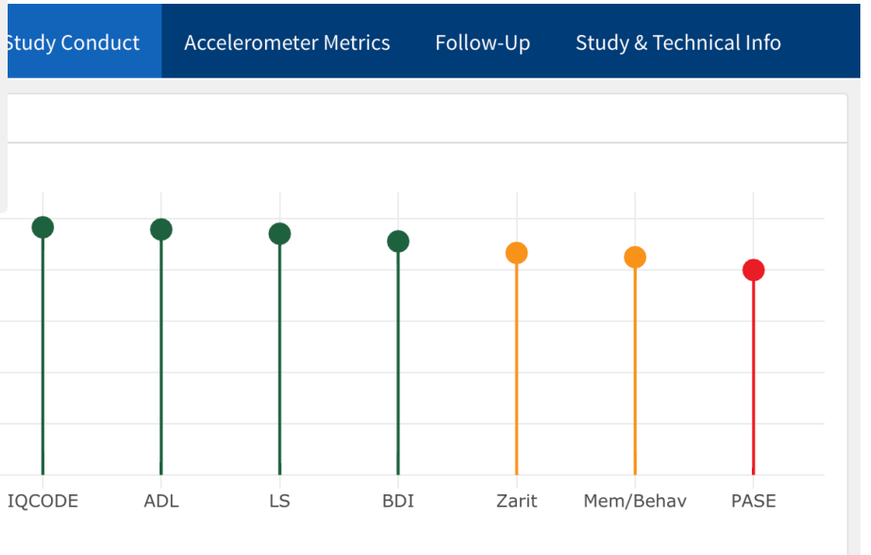
Visual cues **incentivize** targets

Starts **communication** that leads to solutions



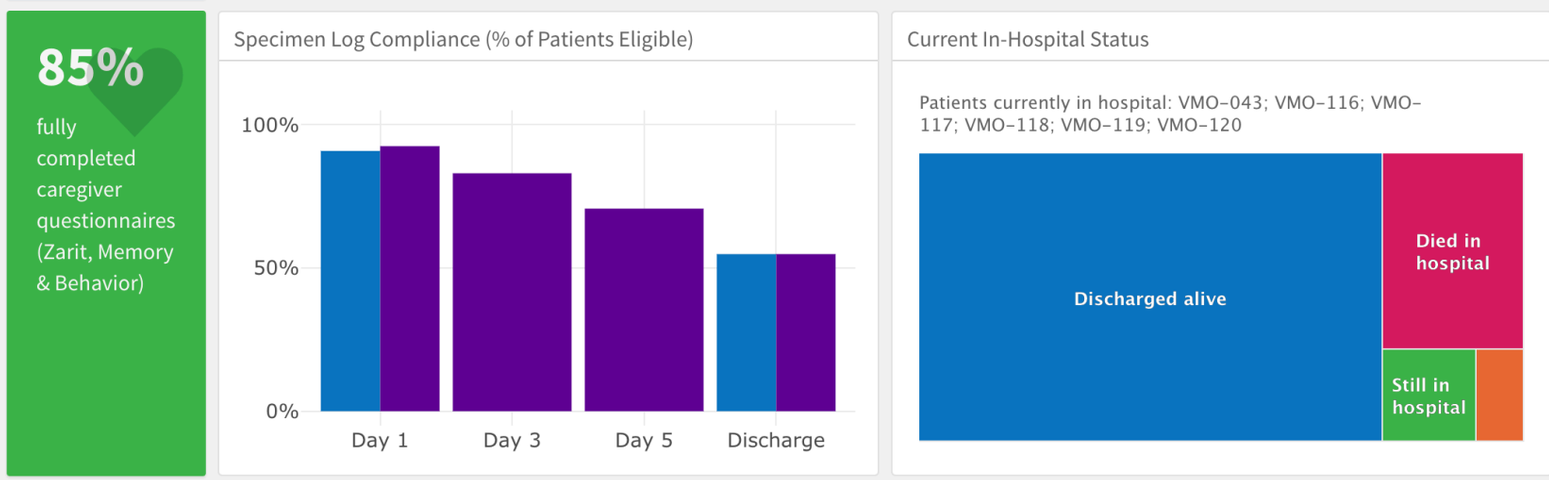
Enrollment tab

- Quick big picture of exclusions, refusals, enrollment
- More detail available via interactive graphics
- Use study logo, color palette for consistent branding



In-hospital tab

- Conditional coloring: quick sense of ability to collect surrogate info
- Lollipops show same for individual instruments
- Specimens collected in blue + purple tubes; using those colors gives quick visual cues
- Patient IDs staff is still following daily = sense of current data collection burden

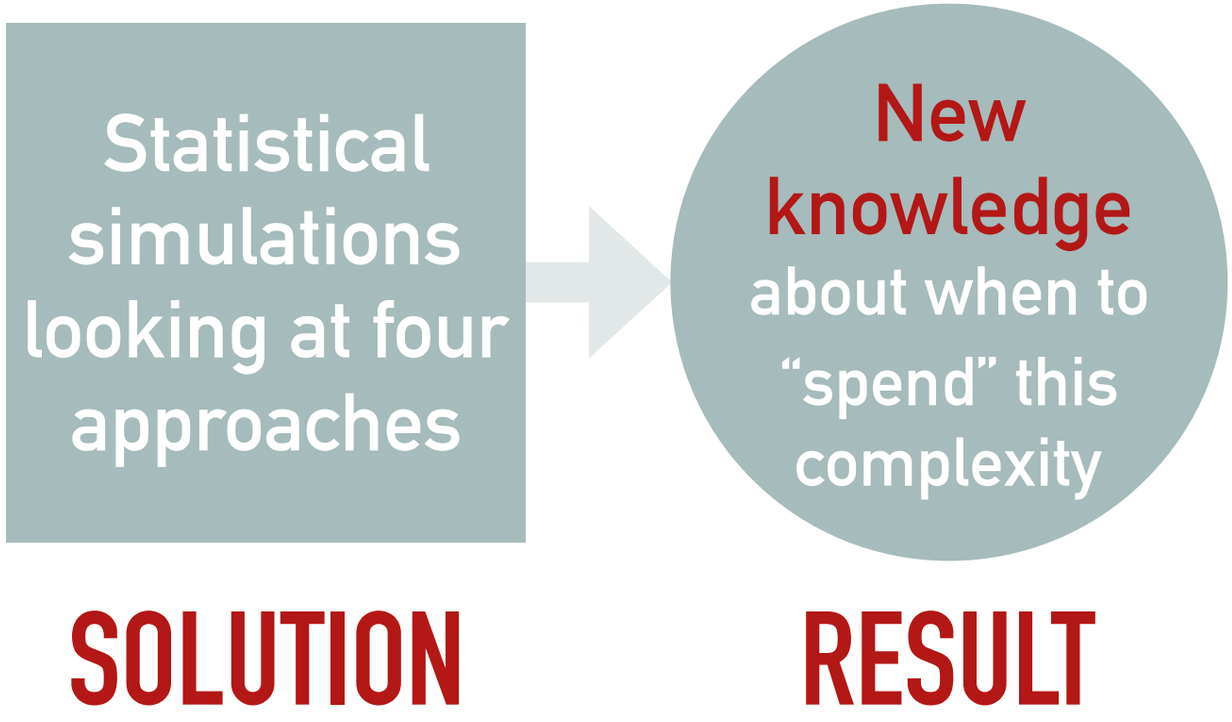


SCENARIO 3:

EMR DATA

SITUATION: USING EMR VS PROSPECTIVELY COLLECTED DATA

- Rich resource with inherent issues, specifically plenty of missingness
- Mental status not assessed as often in trauma unit (subject of this project) as in medical, surgical units; primary exposure relies on these assessments
- We don't know which imputation approach is least likely to bias our results; we assume complex = better, but want to make sure it's worth the resources



Joint Statistical Meetings presentation available at bit.ly/jlt-jsm2018

SCENARIO 4:
MANAGING COMPETING PRIORITIES

SITUATION: ONE GROUP, >10 INVESTIGATORS

WHAT BEST SERVES OUR MISSION?

- As research group has grown, statistical resources have not
- Currently balancing needs of 10 faculty + multiple fellows
- Need communication, prioritization that works

SOLUTIONS
FOR OUR CIRCUMSTANCES

Weekly investigator meetings Slack channel for priority list	"Focus project" vs "ad hoc" weeks	Request form for new projects (aims, $, IRB)	Draft shell of product <u>before</u> analysis begins

RESULTS

Better communication PIs understand priorities, rationale	**Progress toward goals** Major projects move forward, smaller requests covered	Remove logistical **barriers, set expectations** before analysis begins	**Increased efficiency** Complete analysis doesn't wait for someone to take ownership

IN PROGRESS: ONBOARDING DOCS FOR NEW INVESTIGATORS

SCENARIO 5:

CONTINUING EDUCATION

SITUATION: GRAD SCHOOL DOESN'T TEACH US EVERYTHING

- Fields of data science, research, biostatistics constantly evolving; no one can know everything
- Difficult to balance "getting things done" with learning new skills; not all local colleagues have the same level of interest

SOLUTIONS

- Find/create communities that encourage growth
- Create or take advantage of opportunities for training, connections
- Freely share code, knowledge (GitHub, tutorials…)

RESULTS

- **Awareness** of new ideas, research, tools, collaborators
- **Larger toolbox** for solving problems
- Ability to **communicate** with folks from variety of backgrounds
- **Publicity** for methodological, clinical work

THANK YOU!

Talks, writing, CV, links:
jenthompson.me