EHR Data Methodologies in Clinical Research

Using health care data to emulate a target trial when randomized trials are not available

Miguel A. Hernán
Departments of Epidemiology and Biostatistics
Harvard School of Public Health

NIH BD2K Think Tank, December 2014
Focus of this talk

☐ Big Data/EHR for evaluation of interventions
  ■ Comparative effectiveness and safety of clinical and policy interventions
  ■ Causal inference

☐ I will not consider other types of questions
  ■ for example, clinical prediction
Can Big Data Tell Us What Clinical Trials Don’t?

OCT. 3, 2014

When a helicopter rushed a 13-year-old girl showing symptoms suggestive of kidney failure to Stanford’s Packard Children’s Hospital, Jennifer Frankovich was the rheumatologist on call. She and a team of other doctors quickly diagnosed lupus, an autoimmune disease. But as they hurried to treat the girl, Frankovich thought that something about the patient’s particular combination of lupus symptoms — kidney problems, inflamed pancreas and blood vessels — rang a bell. In the past, she’d seen lupus patients with these symptoms develop life-threatening blood clots. Her colleagues in other specialties didn’t think there was cause to give the girl anti-clotting drugs, so Frankovich deferred to them. But she retained her
We need to make decisions NOW

- Treat with A or with B? Treat now or later? When to switch to C?

- A relevant randomized trial would, in principle, answer each comparative effectiveness and safety question
  - Interference/scaling up issues aside
But we rarely have randomized trials

- expensive, untimely, unethical, impractical

- And deferring decisions is not an option
  - no decision is a decision: “Keep status quo for now”

- **Question:** What do we do?
Answer:
We conduct observational studies

- but only because we cannot conduct a randomized trial

- But observational studies are **not** our preferred choice
  - For each observational study, we can imagine a hypothetical randomized trial that we would prefer to conduct
    - If only it were possible
The *target trial*

- An observational study in a **large health care database** can be viewed as an attempt to emulate a hypothetical, nonblinded randomized trial.

- If the observational study succeeds at emulating the target trial, both studies would yield identical effect estimates except for random variability.
Procedure to answer each clinical/policy question:

- **Step #1**
  - Describe the protocol of the target trial

- **Step #2**
  - **Option A**
    - Conduct the target trial
  - **Option B**
    - Use observational (Big) data to *explicitly* emulate the target trial
    - Apply appropriate Big Data analytics
Key elements of the protocol of the target trial

- Eligibility criteria
- Start/End of follow-up
- Strategies/Interventions
  - randomly assigned at start of follow-up
- Outcomes
- Causal contrast(s) of interest
- Analysis plan
The observational study needs to emulate

☐ Eligibility criteria
☐ Start/End of follow-up
☐ Strategies/Interventions
  ■ randomly assigned at start of follow-up
☐ Outcomes
☐ Causal contrast(s) of interest
☐ Analysis plan
Some published examples of an explicit target trial approach

- Hormone therapy and coronary heart disease in postmenopausal women
  - EMRs from the UK / Observational cohort study
- Statins vs. standard of care and risk of coronary heart disease
  - EMRs from the UK
- Individualized strategies to initiate antiretroviral therapy and mortality in HIV-infected patients
  - Health records from Europe and the US
- Individualized strategies for epoetin dosing in hemodialysis patients
  - Claims from USDRS Medicare

- The explicit emulation avoided otherwise common biases and allowed the comparison of complex strategies
Emulation of target trial not straightforward

- For example:
  - There may be insufficient data to characterize individuals eligible for the target trial
  - Unclear whether the outcome ascertainment is accurate
  - etc, etc.

- Use target trial approach to organize discussions about which data are required/missing
“We want to use Big Data as they exist”

☐ First we need to know what exists

☐ Implication

  ■ Only experts users of the data can use them to emulate a target trial
    ☐ Time-varying clinical workflows, idiosyncratic coding practices, software versions...

☐ Also

  ■ Validation studies needed to **quantify** data accuracy
  ■ Cross-datasets comparisons needed to understand coding differences
The target trial will be a compromise

- between the ideal trial we would really like to conduct and the trial we may reasonably emulate using the available data

- The drafting of the protocol of the target trial is typically an iterative process
  - That requires detailed knowledge of the database
Advantages of the target trial approach (I)

- Provides ready access to the application of formal *counterfactual theory and concepts* to Big Data
  - without the need for technical jargon,
- Organizing principle for causal inference methods
  - which implicitly rely on counterfactual reasoning
  - e.g., new user design, active comparators, outcome controls
Advantages of the target trial approach (I)

- Provides ready access to the application of formal **counterfactual theory and concepts** to Big Data
  - without the need for technical jargon,
- Organizing principle for causal inference methods
  - which implicitly rely on counterfactual reasoning
  - e.g., new user design, active comparators, outcome controls
Advantages of the target trial approach (II)

- Facilitates the comparison of complex strategies that are sustained over time and may depend on a patient’s evolving characteristics
  - Dynamic treatment strategies

- Not “treat vs no treat” but rather “when to treat, when to switch, when to monitor” depending on time-varying factors
Advantages of the target trial approach (III)

- Establishes a link between methods for the analysis and reporting of randomized trials and Big Data analytics
  - Observational studies analyzed like randomized trials, and vice versa
Advantages of the target trial approach (IV)

- Naturally leads to analytic approaches that prevent apparent paradoxes and common biases
  - Selection bias related to prevalent users
  - Immortal time bias
  - Birth weight paradox, obesity paradox
  - Etc.
Advantages of the target trial approach (V)

- Facilitates a systematic methodologic evaluation of observational studies
  - which components of the target trial we weren’t able to mimic approximately?
  - which components of the target trial would be problematic even if we were able to conduct a truly randomized trial?
- An approach adopted by the Cochrane Collaboration Risk of Bias Tool for Nonrandomised Studies and the IOM Report on the Safety of Approved Drugs
Advantages of the target trial approach (last)

- If we can influence how data are recorded
  - the target trial approach helps record them

- If we are using data as they exist
  - the target trial approach guides the validation studies and the development and evolution of the Data Model

- The target trial approach allows you to systematically articulate the tradeoffs that you are willing to accept
  - regarding eligibility criteria, interventions confounders, outcomes
Can Big Data Tell Us What Clinical Trials Don’t?

OCT. 3, 2014

When a helicopter rushed a 13-year-old girl showing symptoms suggestive of kidney failure to Stanford’s Packard Children’s Hospital, Jennifer Frankovich was the rheumatologist on call. She and a team of other doctors quickly diagnosed lupus, an autoimmune disease. But as they hurried to treat the girl, Frankovich thought that something about the patient’s particular combination of lupus symptoms — kidney problems, inflamed pancreas and blood vessels — rang a bell. In the past, she’d seen lupus patients with these symptoms develop life-threatening blood clots. Her colleagues in other specialties didn’t think there was cause to give the girl anti-clotting drugs, so Frankovich deferred to them. But she retained her