

Lake Nona: Identifying the right question(s)

IOM ROUNDTABLE ON VALUE & SCIENCE-
DRIVEN HEALTH CARE

Miguel A. Hernán
Harvard School of Public Health

My charge (as stated by IOM staff)

- What kind of questions can you answer with these data?
- Explore limitations and opportunities in a broader sense

- In 10-12 minutes

- That's why I like to attend IOM meetings
 - Each time a greater challenge

I have two simple points

1. Questions and Aims need to be aligned
2. “Small data” determines the question
 - “small data” not my favorite term, by the way

A taxonomy of questions

1. Causal questions

a) Comparative

- e.g., compared with no surgery, does bariatric surgery reduce risk of coronary heart disease?

b) Mechanistic

- e.g., is weight loss a mediator of the effect of bariatric surgery on coronary heart disease?

2. Non-causal questions

a) Descriptive

- e.g., what is the prevalence of obesity?

b) Predictive

- e.g., is obesity associated with a greater risk of coronary heart disease?

“Identifying the right question” of what type?

- Answering each question type requires
 - different data and different tools

- and serves different purposes

For example, comparative causal questions

□ Data / Tools:

- Ideally, randomized experiment
- If experiment not available, observational study

□ Purpose:

- to guide decision making when more than one intervention (including no intervention) are available

- Comparative effectiveness/safety/cost

A taxonomy of causal goals

- a) To change the world
 - Comparative causal questions
- b) To understand the world
 - Mechanistic causal questions
- These goals are non-mutually exclusive
 - But we can classify research projects according to which one predominates
- We choose questions according to goals

What is the goal of a Continuously Learning Health Care System?

- Triple aim (from briefing book)
 - improved care for individuals
 - enhanced health for populations
 - reduced total health care costs
- That is, **to change the world** by choosing the best possible intervention
 - at the individual or population level

- Should we be talking here ONLY about comparative causal inference questions?

A taxonomy of questions

1. Causal questions

- a) Comparative effectiveness/safety/cost
 - e.g., does bariatric surgery reduce risk of coronary heart disease compared with no surgery?
- b) Mechanistic
 - e.g., is weight loss a mediator of the effect of bariatric surgery on coronary heart disease?

2. Non-causal questions

- a) Descriptive
 - e.g., what is the prevalence of obesity?
- b) Predictive
 - e.g., is obesity associated with a greater risk of coronary heart disease?

Examples of questions that not comparative causal questions

- Is fat liver associated with worse outcomes?
 - Is sleep pattern associated with worse outcomes?
 - Psychosocial stressors?
-
- What is the intervention?
 - If no intervention in sight, how can we achieve the triple aim?

My first point

- If we are to achieve the stated aims for the Lake Nona project, we need to think about interventions
 - Comparative causal questions

- Otherwise we are doing something else
 - Possibly very interesting too

A common characteristic of comparative questions

- In principle, they can be answered via a randomized experiment
 - In practice the experiment may not be feasible, ethical, timely, affordable...
- Often the randomized experiment needs to be approximated via an observational study
 - If data on exposure, outcome, confounders

Many observational studies do that

- What's new/better in Lake Nona?

- Or how can one ask better or different questions using Lake Nona data?
 - Using observational data

 - (Lake Nona a good platform for randomized experiments too)

Examples of data collected by existing studies

- Health care organizations collect detailed data on
 - Health care utilization, medical diagnoses
 - Electronic medical records
 - Costs
- Epidemiologic cohort studies collect data on
 - (Self-reported) lifestyle, diet, sleep
 - Environmental exposures
 - Anthropometrics, geocoding
 - Biological samples, including genetics information
- National registries (Scandinavia)
 - Life cycle data, intergenerational data

Lake Nona can collect ALL of these data

- With less measurement error, less missing data
 - Data quality is a KEY consideration for the project

Lake Nona can link ALL of these data

- Integrating high-quality data sources allows
 - Consideration of more detailed interventions
 - Identification of individuals that may benefit more from the intervention
 - Better adjustment for confounding when using observational data

Lake Nona can collect NEW data sources

□ Examples:

- Individual consumer transactions
- Location patterns and other mobile data
- Email / social media
- Netflix, Amazon... usage

□ May suggest new interventions

- Tools to decrease dependence on pain meds
- Decrease media consumption to rebalance sleep habits

The constraints

- Only a few thousand people
 - Common outcomes only: obesity, not cancer
 - Measurable traits: blood pressure
 - Cost
- Answers to comparative questions may be context-specific
 - In fact even the questions themselves may be context-specific
 - Lake Nona does differ from other places
 - residents, health care system, environment, etc.
 - But that's true of many randomized trials too

I have two simple points

1. Questions and Aims need to be aligned
2. “Small data” determines the question