Learning from data to make better decisions in healthcare

A taxonomy of tasks in health data science

Miguel Hernán

HARVARD T.H. CHAN SCHOOL OF PUBLIC HEALTH

Harvard–MIT Health Sciences & Technology

Artificial intelligence (AI) in healthcare

https://en.wikipedia.org/wiki/Artificial_intelligence_in_healthcare

- the use of complex algorithms and software to emulate human cognition in the analysis of complicated medical data
  - the ability for computer algorithms to approximate conclusions without direct human input
- Ahem

Apologies for the letdown but…

AI doesn't really exist

- We are not even close yet
  - "Artificial" yes, "Intelligence" no

- An example of branding success
  - Imagine if we said "curve fitting" rather than "AI"

On a more positive note…

- We have increasingly
  - larger databases
  - better algorithms
  - more powerful computers

- We should be able to implement those **algorithms** on those **computers** to learn something from those **databases**

What can we *actually* learn from data?

- Data scientists often define their work as "gaining insights" or "extracting meaning" from data

- These definitions are way too vague
  - We need a more precise classification of the "insights" and "meaning" that we can learn
  - to think systematically about the types of data, subject-matter knowledge, and analytics that are needed

Starting from the beginning: We can do **3 things** with data

- Therefore, there are 3 types of learning
  - Each one requires different types of data, domain knowledge, and analytics

A Second Chance to Get Causal Inference Right: A Classification of Data Science Tasks

Miguel A. Hernán, John Hsu, and Brian Healy

For much of the recent history of science, learning from data was the academic realm of statistics,[1,2] but in the early 20th century, the founders of modern statistics made a momentous decision about what could and could not be learned from data: They proclaimed that statistics could be

example, the famous "Simpson's paradox" stemmed from a failure to recognize that the choice of data analysis depends on the causal structure of the problem.[6] Mistakes occurred. For example, as a generation of medical researchers and clinicians believed that postmenopausal hormone therapy reduced

discrepancies remained, a unified theory of quantitative causal inference had emerged.[8,9]

We now have a historic opportunity to redefine data analysis in such a way that it naturally accommodates a science-wide framework for causal inference from observational data. A recent influx of

Chance 2019; 32(1):42-49

A taxonomy of data science tasks

Chance 2019; 32(1):42-49

1. Description is using data to provide a **quantitative summary** of certain features of the world
2. Prediction is using data to **map** some features of the world (the **inputs**) to other features (the **outputs**)
3. Counterfactual prediction is using data to predict certain features of the world **if the world had been different**
   - To answer "what if" questions
   - Causal inference is the contrast of counterfactual predictions

Task #1. Description

- Examples of tasks:
  - computing the proportion of individuals with diabetes in a large healthcare database
  - representing social networks in a community
- Examples of analytics
  - elementary calculations (e.g., a mean or a proportion)
  - unsupervised learning algorithms (e.g., cluster analysis)
  - clever data visualizations

Task #2. Prediction (Mapping)

- Often starts with simple tasks
  - association between albumin at admission and death within 1 week in intensive care patients
- then progresses to more complex ones
  - e.g., classification: using hundreds of variables measured at admission to identify patients at high risk of death
- Examples of analytics
  - elementary calculations (e.g., correlation coefficient)
  - supervised learning algorithms (e.g., neural networks)

Task #3. Counterfactual prediction

- Example of task
  - estimation of the mortality rate that would have been observed if all individuals in a population had received cancer screening vs. if they had not received screening
- Examples of analytics
  - elementary calculations in ideal randomized trials (e.g., difference in mortality rates)
  - g-methods in observational studies with treatment-confounder feedback (e.g., the plug-in g-formula)
  - reinforcement learning algorithms

Sciences are defined by their questions, not by their tools

- We define astrophysics as the discipline that learns the composition of the stars
  - not as the discipline that uses the spectroscope

- Similarly, data science is the discipline that uses data for description, prediction, and counterfactual prediction
  - not the discipline that uses, say, deep learning

Data science vs. Data engineering

- Data science benefits from the development of
  - tools for the acquisition, storage, integration, access, and processing of data
  - scalable and parallelizable analytics
- But it is crucial to be explicit about the scientific tasks for which we do that data engineering

Tech companies have excelled at learning from data

- ad placement
- shopping and movie recommendations
- credit rating
- stock trading algorithms
- …

- Data scientists have transferred their expertise to scientific research with biomedical applications

Some early examples of health applications by tech companies

- Google: diagnosis of diabetic retinopathy
  - Data: 120,000 images classified by 54 ophthalmologists
- Microsoft: prediction of pancreatic cancer months before its usual diagnosis
  - Data: online search histories from 3000 users who were later diagnosed with cancer
- Facebook: detection of users who may be suicidal
  - Data: posts and live videos

All these applications have something in common

- They are examples of prediction
  - They map observed inputs (e.g., an image of a human retina) to observed outputs (e.g., a diagnosis of retinopathy)
- They are not examples of counterfactual prediction
  - They do not consider how the world would look like under different courses of action
    - What would happen if we operated on the retina?

Prediction is a natural first step for data science because…

… a successful prediction only requires 3 elements

1. a large dataset with inputs and outputs
2. an algorithm that establishes a mapping between inputs and outputs
3. a metric to assess the performance of the mapping

which means that prediction can be automated

- algorithms improve the mapping without human intervention

Prediction requires little or no domain-specific knowledge about causal structure

- after specifying a well-defined predictive task and acquiring the data, the main ingredient required for success is a good learning algorithm

- This is the key difference between prediction and counterfactual prediction
  - causal inference requires expert causal knowledge to *both* formulate the causal question and to perform the analysis that will provide the answer

Two examples of expert causal knowledge

1. Identification of insufficient adjustment
   - When important confounders are not in the data
     - Screening for colorectal cancer
2. Identification of incorrect adjustment
   - When variables in the data are colliders
     - Maternal smoking and infant mortality

- In both examples, the task is trivial for a human expert but impossible for an "AI" system

Example 1. Does screening colonoscopy prevent all-cause mortality?

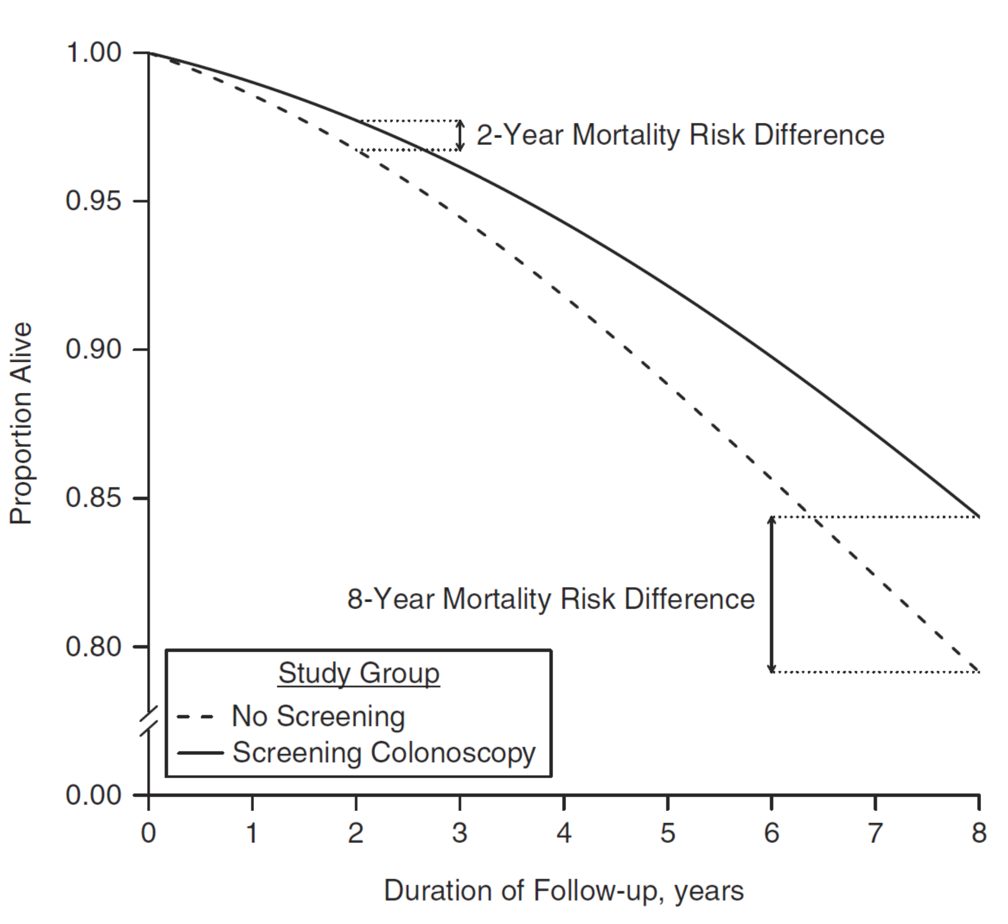
- We have lots of data
  - 430,085 US Medicare beneficiaries
- We have powerful algorithms for confounder identification and adjustment
  - High-dimensional propensity score
- We have good computers

- Do we get a reasonable answer?

No, our effect estimates make no sense

Garcia-Albeniz et al. Am J Epidemiol 2019; 188(10): 1764-1767

- 8-year mortality risk difference
  - −6.3 percentage points
- But only ~2% of people die of colorectal cancer!



Important confounders were not in the data

- Smoking, physical activity… and measures of health consciousness

- No clever algorithm could adjust for them
- Even if we had data on 10 billion people
- And the fastest supercomputer on Earth

- But our learning algorithm could not know that

Example 2. Does maternal smoking affect infant mortality?

- Pregnant women who do and do not smoke differ in many characteristics that affect the risk of infant mortality
  - alcohol drinking, diet, access to adequate prenatal care…

- Suppose all confounders are in the data
  - And therefore can be successfully adjusted for

Problem

- Confounding factors are associated with both maternal smoking and infant mortality

- But not all factors associated with maternal smoking and infant mortality are confounders
  - Adjustment for some of these factors introduces bias

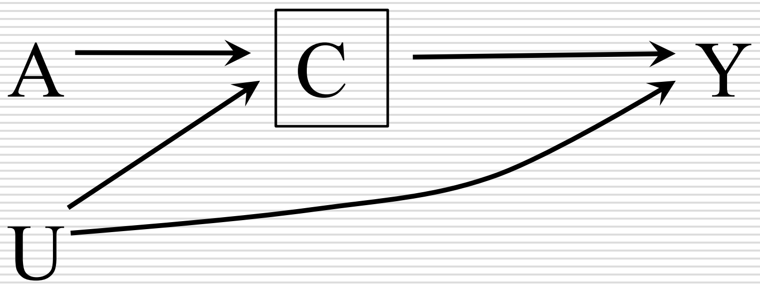
- How can a a learning algorithm know this?

Take the variable “birth weight”

- strongly associated with both maternal smoking and infant mortality
- Adjustment for birthweight induces bias
  - because birthweight is a risk factor that is itself affected by maternal smoking
- Adjustment for birth weight results in the “birthweight paradox”
  - low birth weight babies from mothers who smoked have a lower mortality than those from mothers who did not smoke during pregnancy

The birth weight paradox

Hernandez-Diaz et al. *Am J Epidemiol* 2006; 164(11): 1115-1120



- A: Cigarette smoking
- U: birth defects, intrauterine malnutrition (unmeasured)
- C: low birth weight
- Y: infant mortality

An algorithm that relies on associations in the data can lead to wrong causal conclusions

- Because it will miss features that should be adjusted for but are not in the database
  - A human expert can readily identify the missing variables

- Because it will adjust for features that increase bias and lead to the incorrect causal conclusion
  - A human expert can readily identify these variables

Causal inference cannot be reduced to data analysis

- Except under the extreme assumption that the data contain
  - No unmeasured confounders
  - No colliders
  - No instruments

- Data analysis competitions work for prediction but not for counterfactual prediction (causal inference)

Comment: Spherical Cows in a Vacuum: Data Analysis Competitions for Causal Inference

Miguel A. Hernán *Statistical Science* 2019; 34: 69-71

Researchers using causal inference competitions for practical advice are like the farmer who asked the local university for assistance to increase the milk production of his cows.

A sophisticated theoretician wrote back:
"I have the solution, but it works only in the case of a spherical cow in a vacuum."

Answering causal questions requires data, analytics, *and* subject-matter knowledge

- Common source of confusion
  - A failure to grasp the different role of subject-matter causal knowledge in prediction and causal inference

- The data are insufficient to evaluate the performance of causal inference tasks
  - Splitting data into training and test subsets doesn’t help

But we have known all of this for a long time

- Healthcare databases have been used for decades
  - By epidemiologists
    - Hershel Jick used EHRs in the 1970s and 1980s
- Biobanks have existed for decades
  - They were called epidemiologic cohorts
    - Framingham Heart Study since the 1950s (<10k)
    - Nurses' Health Study since the 1970s (>100k)
    - Many others
- Learning algorithms have been used for decades
  - We called them data analysis

What’s new then?

- Increased computational power
  - Algorithms developed in the 1980s and 1990s (e.g., neural networks) can now be realistically implemented
  - Larger databases can be handled
- Financial support from tech companies
  - Amazon, Google, Apple, Microsoft…
- Sexier terminology
  - “Careful exposure measurement” is now “deep phenotyping”
  - “Curve fitting” is now “a form of AI”

Computer scientists are reinventing the wheel. This is wasteful

- Lots of work to rediscover well-known facts
  - healthcare data are complicated and messy
  - getting high-quality data is important
  - deep phenotyping (diet, metabolomics...) requires epidemiological cohorts
- Creation of a new terminology for well-known concepts
  - redundant terminology hampers communication across disciplines

Computer scientists are reinventing the wheel. This is wasteful

- Mistakes are repeated
  - Cautionary tales have to be re-generated

- **Computer scientists are distracted from what they do best**
  - and we do need computer scientists

Why the confusion then?
Because of extrapolation from simple settings

- Distinguishing between prediction and causal inference is irrelevant
  - when all expert knowledge can be incorporated into the algorithms
- A purely predictive algorithm that learns
  - to play Go can perfectly predict the counterfactual state of the game under different moves
  - to drive a car can accurately predict the counterfactual state of the car if, say, the brakes are not operated

Take Go: a game mastered by an algorithm “without human knowledge”

- When making a move, the algorithm has access to all information that matters:
  - game rules
  - current board position
  - future outcomes
  - a reinforcement learning algorithm can collect an arbitrary amount of “experimental” data
- Healthcare questions are not like that

In these relatively simple settings

- An algorithm can predict the behavior of the entire system under a hypothetical intervention
  - because these systems are governed by a set of known game rules (in the case of games like Go) or physical laws with some stochastic components (in the case of engineering applications like self-driving cars)
- Prediction implies counterfactual prediction
- Healthcare questions are not like that

Implications for decision making

- A premise of data science is that it can help make better decisions
  - the ability of data science to improve decision making is predicated on the basis of its success at prediction.
- However, predictive algorithms do not directly guide decision making
  - They alert about the need to make a decision, but do not recommend which one

Example: Prediction of 5-year mortality among patients with heart failure

- Identifying patients with bad prognosis is very different from identifying the best course of action to prevent mortality
  - prior hospitalization may predict death, but nobody suggest stops hospitalizing people to prevent death
- An algorithm that excels at identifying high-risk patients is agnostic about how to reduce death risk
  - and may not be transportable across settings

Learning from data and decision making in healthcare

- Main selling point of AI/machine learning
  - Better decision making by finding out "what works for whom"
    - Personalized treatment
  - This is causal inference
- What most AI people do
  - Diagnostic classification, identifying new biomarkers, finding out high-risk patients…
  - This is not causal inference

A data science (or AI or machine learning) that embraces causal inference must

1. develop methods for the integration of sophisticated analytics with domain expertise
2. acknowledge that, unlike for prediction, the assessment of the validity of causal inferences cannot be exclusively data-driven
   - because the validity of causal inferences also depends on the adequacy of expert causal knowledge

Remember: No AI will be worthy of the name without causal inference

- A hallmark of intelligence is the ability to predict *counterfactually* how the world would change under different actions

Thank you.

- For more info:
  - Causal Inference book
    - Free online, google "causal inference book"
  - Causal diagrams course
    - Free online, google "edX causal diagrams"
  - Twitter: @_MiguelHernan