Using Data to Advance Knowledge in Healthcare: Historical Challenges and the Small Data Problem

PAUL GERRARD, MD

UNIVERSITY OF SOUTH CAROLINA BIG DATA IN HEALTH SCIENCE CONFERENCE

2/7/2020

Disclosures

- I am an employee of Palmetto GBA
- All opinions expressed are strictly my own

Healthcare has not really seen much benefit from big data

....adoption of big data analysis in healthcare has lagged behind other industries ... In the meantime, 80 percent of executives from financial services, insurance, media, entertainment, manufacturing, and logistics companies surveyed report their investments in big data processing as "successful," and more than one in five declare their big data initiatives have been "transformational" for their firms.

Catalyst, N.E.J.M., 2018. Healthcare big data and the promise of value-based care. *NEJM Catalyst*, *4*(1).

A brief detour in history

Early Use – John Graunt (1620-1674)

- A work titled: *Observations on the Bills of Mortality*
- Used data to develop tables of the causes of death 1647-1659 in London
- Data came from churches
- He attributes the motivation to start data collection to the plague
- You can read the whole text (I wouldn’t recommend it) here: http://www.edstephan.org/Graunt/bills.html

- No regression, no t-tests, just counting and tabulation
- Data analysis in healthcare today is generally pretty similar to what John Gaunt did
  - Using relatively small numbers
  - Tabulation and counting is as sophisticated as many analyses get

What does “big data” mean?

- “Big data” is a term coined by Roger Magoulas of O’Reilly media in 2005
- Originally the term was used to refer to data sets that were very difficult to manage due to size and complexity
- Tools for handling large data sets were developed to allow distributed computing and take advantage of immensely large data sets
  - Hadoop distributed file system
  - Map Reduce programming model
  - Apache Spark
  - CERN Large Hadron Collider Computing Grid
  - NVIDIA’s CUDA
- Do all of these tools have a role in healthcare data? Do any of them?
- With the development of better technology and optimization algorithms, data size and complexity have started to be seen as sources of opportunity rather than difficulty
- The meaning of “big data” has morphed, such that it is now a term used to describe predictive analytics more generally

Healthcare has a small data problem

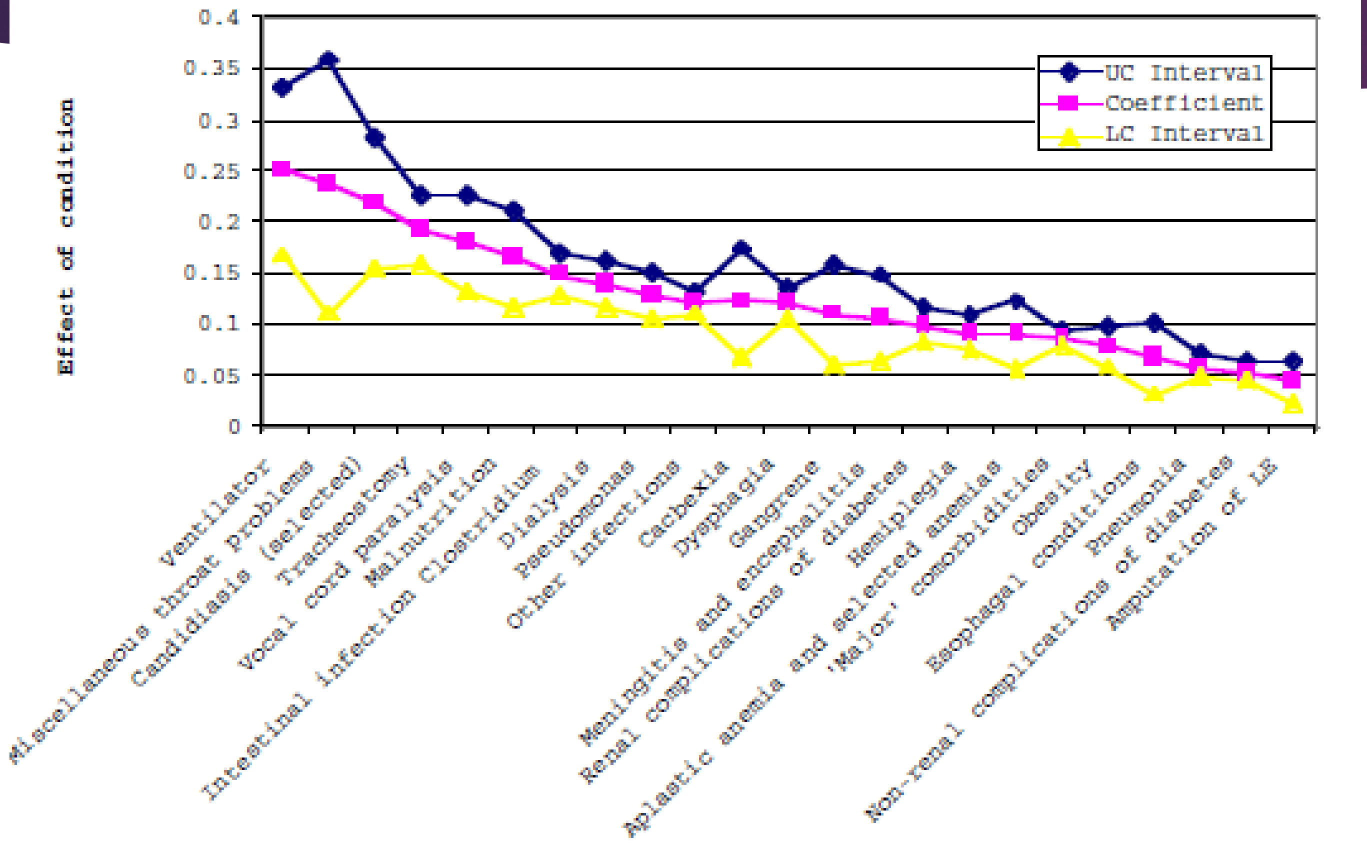
- Immensely large data sets open up many opportunities
  - New tools to make revolutionary big data analysis practical (e.g. CUDA)
- Such data sets are almost non-existent in healthcare
  - Data in general is tough to get in healthcare
  - If you can get a data set, it is likely very limited
  - Data sets are scattered
    - Same patient in different electronic health records
    - Data for payment, textual medical record data, and imagining data may all be in unique locations within a healthcare system (but all within the system)
    - Genomic data is a new data element that is usually held outside of the healthcare system
- This brings two challenges:
  1. Figuring out how to get access to all of these different data elements
  2. Figuring out how to meaningfully integrate them to be usable

Example of small data problem in healthcare

- In 2002 Rehabilitation Hospitals (IRFs) switched to a prospective payment system
- Payment would be driven by characteristics of the patient at admission
- Characteristics of patients captured in a CMG (case mix group) and comorbidity tier

- The Rand Corporation was hired to develop the PPS

Impact of comorbidities on costs



“Experts” identify comorbidities that they think matter

Run linear regression to identify the impact of these comorbidities on cost controlling for RIC

From Rand Report: Analyses for the Initial Implementation of the Inpatient Rehabilitation Facility Prospective Payment System

Ventilator dependence is a low volume diagnosis – see comments in response to CMS's proposed new payment model

"One commenter suggested that spinal cord injury (SCI) patients who are ventilator-dependent should have their own CMG and an associated payment…under the proposed CMGs, an SCI ventilator-dependent patient would always result in an outlier payment… while there is not a large number of these patients, the outlier payment could result in a large financial loss to providers."

CMS's Response:

"We are not including a separate CMG for ventilator-dependent, spinal cord injury patients in this final rule. We will consider analyzing this group of patients for future refinements. **Our current CMGs are based on historical data. In order to develop a separate CMG, we need to have data on a sufficient number of cases to develop coherent groups.** As the commenter noted**, the data that RAND analyzed did not have a sufficiently large number of these patients.** The cost of caring for ventilator-dependent spinal cord injury patients is reflected in the relative weights for the CMGs in which these cases fall. Ventilator-dependent spinal cord injury cases will be classified to comorbidity tier 1. We grouped these types of cases only with other very expensive spinal cord injury patients, and the relative weights set forth in this final rule reflect the average cost for these cases. Therefore, we believe that the standard IRF prospective payment plus the outlier payment (which addresses the marginal cost of care beyond the applicable threshold) will pay adequately for these cases. **It is certainly possible that, for a given case, the total payment for the case might be lower than the cost for the case, but for other cases, the total payment might be higher than costs**."

– Federal Register Volume 66, Issue 152 (August 7, 2001)

Homework

- Look at CMS claims based measures
- For claims based quality measures CMS uses a predicted to expected ratio rather than an intutitive observed to expected ratio
- Criticized as being too complicated and difficult to understand by the Committee of Presidents of Statistical Societies
- Try to figure out what CMS did and why they did it? (Hint – it relates to the problem of “low information” and a need for stabilization)

This is mostly a sociology talk – We have not benefited from big data in healthcare. Why?

It’s not a problem of math, computer science, or existing technology

- Humans / potential patients are doing things that represent potentially useful health data
- Some things are tough to get data on, but there are a lot of things where the data collection is easily achievable with existing technology
- Lots of data analysis tools and computational tools exist

- If we as a individuals and a society simply choose to collect the data that is collectable and use the computational tools available, we can advance our knowledge in healthcare in ways we never have.
- Society has not yet chosen to do this
  - Do we even realize that we have not chosen to?

What can we do to make learning from data possible in healthcare?

- We have to pull down barriers to accomplishing this
- What are those barriers?

Why is data unavailable? Two classes of barriers

- Cultural barriers within healthcare institutions
  - Territoriality, Egos, and personalities
  - Communication problems / misunderstandings
    - Lack of institutional appreciation
    - Perceived costs

- Structural and technical issues
  - Until recently, healthcare data was stored in paper records
  - Electronic medical records are specialized databases which cannot be easily queried

  - HIPAA – law that controls sharing of healthcare information to protect patient privacy
  - Patient concerns – how much data do you want collected on you, and how much do you want to share?
    - This is already a visible concern with Google and social networks where a law like HIPAA does not exist.

Cultural barriers -

- Do we even understand data in healthcare?
- Do we understand basic mathematical concepts?

Conversation from real life

- Large and prestigious academic hospital in New England has a patient safety committee
- Director of nursing says to the committee:
  - While the number of falls per patient day is higher this year, things have gotten a little better, because the total number of falls has declined this year.

- What is the problem with this statement? (Hint: hospital volume is not constant from year to year)
- These kinds of conversations that are happening over data around the country

Example from academic world in medicine

- Manuscript submitted looking at readmissions in stroke patients from rehabilitation hospitals
- Sample contained 70% of the population of interest with no missing data. We have roughly 1 million patients in the sample.
- Comment from a reviewer at one of the most prestigious journals on stroke:
  - With only 70% you have a missing data problem. What did you do about the 30% missing data?
- Another reviewer commented that with such a large data set, “big data” we should be able to include other variables in our model that we did not, and he / she listed a series of variables that the data set did not contain
  - Very common misunderstanding that fails to recognize that you only have the data that someone has chosen to collect

Doctors have an “uncomfortable” relationship with math

- Study evaluating the ability of physicians to calculate a positive predictive value (PPV) of a diagnostic test:

  Our results show that the majority of respondents in this single-hospital study could not assess PPV in the described scenario....

  Statistical reasoning was recognized to be an important clinical skill over 35 years ago,[1-3] and notable initiatives... have developed recommendations to improve the next generation of medical education.[4,5] Our results suggest that these efforts, while laudable, could benefit from increased focus on statistical inference.

  - From: Manrai, A.K., Bhatia, G., Strymish, J., Kohane, I.S. and Jain, S.H., 2014. Medicine’s uncomfortable relationship with math: calculating positive predictive value. *JAMA internal medicine*, *174*(6), pp.991-993.
- What about other healthcare providers?
- Psychologists (particularly neuropsychologists) have a much more robust tradition of quantitative training

What do psychometricians think about use of statistics in psychology?

...even though psychometric modeling has seen rapid and substantial developments in the past century, psychometrics, as a discipline, has not succeeded in penetrating mainstream psychological testing to an appreciable degree....

The psychometric routines commonly followed by psychologists working in 2006 do not differ all that much from those of the previous generations. .... As such, contemporary test analysis bears an uncanny resemblance to the psychometric state of the art as it existed in the 1950s...

This problem is compounded by the research standards that are currently accepted in psychology. Even though the research topic of psychology—human behavior and the mental processes that underlie it—is perhaps the most complicated ever faced by a science, the contents of scientific papers that deal with it are required to be below a certain standard of difficulty. I have seen at least one case where a manuscript that used psychometric modeling was rejected by a major journal because, according to the editor, it was too difficult for the journal's audience since it contained some basic matrix algebra ....

Borsboom, D., 2006. The attack of the psychometricians. *Psychometrika*, *71*(3), p.425.

Had things changed much over the next 7 years?

- Mark Wilson wrote in 2013 that while psychometrics is advancing based on the work of psychometricians, these advancements are not making their way into psychology and sociology research or education.

One of the consequences of this is that many new psychometric models …are beyond the reach of typical substantive researchers. Hence, substantive researchers and users (e.g., research psychologists and people in the production divisions of testing companies) tend towards the following:

> *(a) they maintain "standard measurement approaches" that are not well-aligned with more recent developments in psychometrics …, and/or*
>
> *(b) they develop approaches that derive from alternative perspectives to psychometrics(e.g., non-quantitative perspectives, item-focused approaches).*

Wilson, M., 2013. Seeking a balance between the statistical and scientific elements in psychometrics. *Psychometrika*, *78*(2), pp.211-236.

Summary

- There is a cultural barrier to adoption of data analysis in healthcare, which is related to perceptions and understanding
  - This exists even among clinicians, clinical leaders, and clinical researchers,
  - This affects multiple generations
  - This is widespread among different types of doctoral-level licensed healthcare professionals

- Nobody seems to argue that better understanding of quantitative methods is important in healthcare, but it is not clear that this is happening in spite of awareness of this need

Lets move on to the structural barriers

Available data sets in Healthcare don't allow us to address large hypothesis spaces

- Data depth and volume
- Hypothesis space size

- The key to practical data science is creating a hypothesis space that addresses the question you have, and which can be estimated using the data available.
- If you want to use tools that rely on massive amounts of data, you need to collect massive amounts of data.
- Data costs money
  - Data that is already being collected lends itself best to large volumes
  - Specially collected data will be extremely costly to collect in large volumes

Recipe for using big data in healthcare

1. Extract data being collected and curate it with a minimum quality standard

    Optional – collect additional data not yet being collected

2. Aggregate available data on patients from all possible sources

3. Develop **well-specified** models with a limited number of parameters to reduce hypothesis space or improve the information content of existing data

- Entities able to do all 3 of these will be the first movers in healthcare big data revolution
- The ability to analyze the data is not included in the recipe

Let’s talk about data infrastructure of existing healthcare data that is awaiting analysis

What’s needed to use big data tools?

- Data

- Where does the data reside?
- What is the data quality?
- How can we make the most use of what we have?

Data sources

- Clinical healthcare data
  - Collected on every patient every time they see a doctor or go to the hospital
- Research databases
  - Specifically collected for research, usually on people in research studies
  - Generally much smaller data set than clinical data
- Lifestyle and fitness data
  - Wearable devices / fitness trackers
  - Location?
  - Anything that could possibly be informative that is not in the top 2 data sources

Clinical data – This is collected as a part of routine clinical care. No extra resources are devoted to data collection.

- Electronic medical records
  - Physician notes
  - Vital signs
  - Diagnostic testing reports
  - Medication lists
  - Etc…
- Imagining data
  - Actual pixel-level (or voxel) data
- Laboratory data
- Administrative / billing data
  - Encounter dates
  - Diagnoses
  - Treatments
  - Etc….
- Some of this data is not entered in real time, but after all care decisions using it have already been made

Real world evidence

- Includes data gathered from patients outside of a research setting
- Immense interest in this

- 21st Century Cures Act signed into law in 2016 amended the Food, Drug, and Cosmetic Act as follows:

  §355g. Utilizing real world evidence

  (a) In general

  The Secretary shall establish a program to evaluate the potential use of real world evidence-

  (1) to help to support the approval of a new indication for a drug approved under section 355(c) of this title; and

  (2) to help to support or satisfy post approval study requirements.

  ...

- The FDA has been directed by law to consider real world evidence

At least two separate applications of clinical data in healthcare

- Research – The use of data (including clinical data) to learn generalizable principles that can be applied broadly and are typically published in academic press
  - Sometimes clinical data is used
  - Sometimes specially obtained research data is used
- Clinical care / Quality data
  - The use of data to drive care locally
  - The use of care to benchmark performance – tell you about the past, though not necessarily in a way that is generalizable to the future

Let’s not lose sight of culture?

- Where is the vast majority of data generated?
  - Clinical care
- Who has the most interest and experience in using data?
  - Researchers
- Conventional healthcare data is almost entirely generated in clinical care
- What about institution-level healthcare quality measures, clinical policies, and care models?
  - Generally put together by clinicians

- Is there alignment between data infrastructure and culture of data use?
- The individuals and groups with a background to use data are different from the individuals and groups collecting most conventional healthcare data

Data volume

- Machine learning describes a huge array of techniques
- Nearly all techniques are at some level concerned with estimated parameters in a model
- Shallow techniques estimate a very limited number of parameters
  - E.g. linear regression might estimate as few as 2 parameters, slope and intercept.
- Deep learning estimates many parameters
  - A neural network may require estimation of thousands of parameters
- More parameters usually requires more data

Why are companies like Google and Facebook able to use data hungry tools?

- Have records from thousands of clicks on each user
- Hundreds or thousands of days of data for each user
- Time data
- May have data on other websites visited
- All digital data (there unlikely to be corrupted)

- Data in healthcare does not look like data collected by these tech giants

What does data look like in healthcare

- At least conventional healthcare

Sources of large data in healthcare

- ICU patients
  - Potentially very deep data set, but small volume of patients
  - Constant monitoring of the electrical signals in the heart
  - Constant monitoring of blood oxygen levels
  - Frequent (maybe even constant) blood pressure measurements
- Claims data sets
  - May have massive volume of patients, but likely very limited in depth / detail
  - Demographic information
  - Medical procedure codes
  - Dates
  - Diagnoses by ICD-10 code

- Does this data have quality problems?
- These can be large data sets (at least for healthcafe), but are we even collecting important information
- Does any of this information necessarily tell you whether the patient is comatose or awake and speaking? No
- Is any of this data even a good measure of health or useful in predicting adverse outcomes in the first place?

Example of limitation

- Heart failure is a common diagnosis; high blood pressure is a common cause
- Usual symptom is shortness of breath
- We as doctors are trained to think about diagnosis
- ICD-10 code for heart failure due to high blood pressure
  - I11.0

- Are these two patients the same?
  - One has shortness of breath only when doing more than average activity
  - One has shortness of breath all the time, even at rest
- Both have the same diagnosis
- Both would get the same ICD-10 code (even the medical record may not tell the difference)
- Evidence-based treatment in terms of medications might even be the same

What's missing?

- Severity measures of disease
- Conventional healthcare data is much better at capturing the diagnosis than the severity of the diagnosis. Even physician notes often tend not to capture severity information.

- What was I alluding to in the heart failure example?
- Ability to physically function as a severity measure.

- Well…there is at least one place that functional status data is available in administrative data sets.

Research has shown that functional status is a very good predictor of

C-Statistics of 30-Day Readmission Risk Prediction Models by Impairment Category

Impair Group	Demographic-Comorbidity Models			Function Only Models	Function Plus Models		
	Demo + Elixhauser	Demo + Deyo-Charlson	Demo + CMS Tiers	FIM	FIM + Demo + Elixhauser	FIM + Demo + Deyo-Charlson	FIM + Demo + CMS Tiers
CVA	0.59 (0.58, 0.59)	0.56 (0.56, 0.56)	0.58 (0.58, 0.58)	0.68 (0.68, 0.68)	0.69 (0.69, 0.69)	0.68 (0.68, 0.68)	0.68 (0.68, 0.69)
Brain	0.59 (0.59, 0.59)	0.58 (0.57, 0.58)	0.58 (0.58, 0.58)	0.65 (0.65, 0.66)	0.67 (0.67, 0.68)	0.67 (0.66, 0.67)	0.66 (0.66, 0.67)
Neuro	0.62 (0.61, 0.62)	0.59 (0.59, 0.59)	0.60 (0.60, 0.60)	0.65 (0.64, 0.65)	0.68 (0.68, 0.69)	0.67 (0.67, 0.67)	0.66 (0.66, 0.67)
SCI	0.61 (0.61, 0.62)	0.59 (0.59, 0.60)	0.60 (0.60, 0.60)	0.65 (0.65, 0.66)	0.68 (0.68, 0.69)	0.68 (0.67, 0.68)	0.67 (0.67, 0.67)
Amp	0.60 (0.59, 0.60)	0.58 (0.57, 0.58)	0.58 (0.58, 0.58)	0.67 (0.67, 0.67)	0.69 (0.68, 0.69)	0.68 (0.67, 0.68)	0.68 (0.67, 0.68)
Arth	0.60 (0.59, 0.60)	0.57 (0.56, 0.58)	0.58 (0.57, 0.59)	0.65 (0.64, 0.66)	0.67 (0.66, 0.68)	0.66 (0.65, 0.67)	0.66 (0.65, 0.67)
Pain	0.60 (0.59, 0.61)	0.58 (0.57, 0.59)	0.59 (0.58, 0.60)	0.67 (0.66, 0.68)	0.68 (0.67, 0.69)	0.68 (0.67, 0.69)	0.68 (0.67, 0.69)
Ortho	0.63 (0.63, 0.63)	0.62 (0.61, 0.62)	0.62 (0.62, 0.62)	0.70 (0.70, 0.70)	0.72 (0.72, 0.72)	0.71 (0.71, 0.72)	0.71 (0.71, 0.71)
Card	0.56 (0.55, 0.56)	0.53 (0.53, 0.54)	0.54 (0.54, 0.55)	0.64 (0.64, 0.64)	0.65 (0.65, 0.65)	0.64 (0.64, 0.65)	0.64 (0.64, 0.65)
Pulm	0.57 (0.56, 0.58)	0.55 (0.54, 0.56)	0.55 (0.55, 0.56)	0.66 (0.66, 0.67)	0.68 (0.67, 0.68)	0.67 (0.66, 0.67)	0.67 (0.66, 0.67)
Burns	0.61 (0.58, 0.63)	0.56 (0.53, 0.58)	0.58 (0.55, 0.60)	0.65 (0.63, 0.67)	0.67 (0.65, 0.70)	0.66 (0.64, 0.68)	0.66 (0.64, 0.69)
Cong	0.60 (0.56, 0.64)	0.56 (0.52, 0.60)	0.60 (0.56, 0.64)	0.64 (0.60, 0.68)	0.69 (0.65, 0.72)	0.66 (0.62, 0.70)	0.66 (0.63, 0.70)
Other	0.60 (0.59, 0.61)	0.56 (0.55, 0.57)	0.57 (0.56, 0.58)	0.66 (0.65, 0.66)	0.69 (0.68, 0.70)	0.67 (0.66, 0.68)	0.67 (0.66, 0.68)
MT	0.58 (0.58, 0.59)	0.57 (0.56, 0.58)	0.60 (0.59, 0.60)	0.64 (0.64, 0.65)	0.66 (0.66, 0.67)	0.66 (0.65, 0.66)	0.66 (0.65, 0.67)
Deb	0.59 (0.58, 0.59)	0.56 (0.55, 0.56)	0.57 (0.57, 0.57)	0.65 (0.64, 0.65)	0.67 (0.67, 0.67)	0.66 (0.65, 0.66)	0.66 (0.65, 0.66)
MC	0.57 (0.57, 0.58)	0.55 (0.54, 0.55)	0.56 (0.56, 0.57)	0.65 (0.64, 0.65)	0.67 (0.66, 0.67)	0.65 (0.65, 0.66)	0.65 (0.65, 0.66)
Min	0.56	0.53	0.54	0.64	0.65	0.64	0.64
Max	0.63	0.62	0.62	0.70	0.72	0.71	0.71

Shih, S.L., Zafonte, R., Bates, D.W., Gerrard, P., Goldstein, R., Mix, J., Niewczyk, P., Greysen, S.R., Kazis, L., Ryan, C.M. and Schneider, J.C., 2016. Functional status outperforms comorbidities as a predictor of 30-day acute care readmissions in the inpatient rehabilitation population. *Journal of the American Medical Directors Association*, *17*(10), pp.921-926.

Evidence suggests that functional status data is a good measure of health

- Who collects good functional status data?
- What do we want this data to capture?
  - Massive range of intensity
  - Frequency
  - Quantitative data on a absolute scale

- Medical records on patient's don't
- Administrative data doesn't
  - FIM is a 1-7 ordinal scale on 18 items (range of 18-126)
  - Most community dwelling people will score 126
- What about wearable devices?
  - Garmin
  - Fitbit
  - Apple
  - Samsung
  - Fossil
  - Etc.....
- Is it possible that Garmin knows more about your health than your doctor?
- Not to say that any of these companies are aggregating and sharing your data, but the data is there

Let’s just assume the quality is good. What kind of data is where?

- These are all different databases and disaggregated
- Electronic medical records
  - Physician notes
  - Vital signs
  - Diagnostic testing reports
  - Medication lists
  - Etc…
- Imagining data
  - Actual pixel-level (or voxel) data
- Laboratory data
- Administrative / billing data
  - Encounter dates
  - Diagnoses
  - Treatments
  - Etc….

Disaggregation has important consequences

- Many labs doing genomic testing exist.
  - They have lots of genetic data
  - The labs have very limited if any clinical / outcome data
  - Someone has to merge multiple data sources to have a useful database to mine
- Novel image analysis techniques are coming online
  - Same problem as the labs
  - How do they do meaningful radiomics without this data
- “Multi-Omics”
  - Exciting area combining genetic and radiologic features
  - Will require aggregating multiple data sets to develop

Real world data to develop real world evidence

- Use of existing clinical data to drive innovation

- Companies can use it to make new discoveries
- Companies can use it to support FDA approval – see previous slide
- Companies can use it to support payment for new services by third-party payers

- Hopefully good for patients
- Hopefully good for profits

- If clinical data is so disaggregated, how do we get it?

Solutions

- Two broad categories of barriers addressed
  - Structural / technical barriers
  - Cultural barriers

Let’s look at one company that addresses structural challenges

Flatiron Health

genomeweb

Business & Policy | Technology | Research | Diagnostics | Disease Areas

Roche Completes $1.9B Flatiron Health Acquisition

Apr 06, 2018 | staff reporter

NEW YORK (GenomeWeb) – Roche said today that it has completed a previously announced $1.9 billion acquisition of Flatiron Health, a provider of electronic health record software with a focus on oncology.

Flatiron curates and develops real-world evidence for cancer research. It has a network of community oncology practices and academic medical centers across the

What does Flatiron Health do?

- Offers the OncoCloud Suite to community oncology practices
  - Electronic Medical Record
  - Billing system
  - Provides the practice with data on quality, financial performance, and operational efficiency
- Provides real world evidence to life sciences companies

- What are they in the business of?
  - They provide an EMR, and also…
  - Data collection, extraction, and aggregation

Roche is a massive drug company

- Roche does a lot of work in the cancer space
- Many new cancer drugs focus on specific genetic mutations found in cancer.

- How might Roche they take advantage of this data?

They bought a genomics company

genomeweb

E-mail | Pas

Show p

Business & Policy | Technology | Research | Diagnostics | Disease Areas | Applied Markets | Reso

Roche to Acquire Outstanding Shares of Foundation Medicine for $2.4B

Jun 19, 2018 | staff reporter

NEW YORK (GenomeWeb) – Roche will acquire the remainder of Foundation Medicine for $2.4 billion in cash, the companies said today.

Roche already owned about 57 percent of Foundation's outstanding common stock, following a tender offer in 2015.

Under a definitive merger agreement, Roche will acquire all outstanding shares of Foundation Medicine that Roche or its affiliates don't already own at a price of $137 per share in cash, a 29 percent premium over Foundation's closing price on Monday and a premium of 47 percent and 68 percent over its 30-day and 90-day volume weighted average share price, respectively, on June 18.

What is Foundation Medicine

- A genetics testing company for cancer
- Has the FDA-approved FoundationOne Companion Diagnostic
- One of the most commonly ordered cancer genetic tests

- In 2014 Foundation Medicine and Flatiron Health announced a collaboration to combine genomic and clinical data
  - i.e. aggregation of clinical and genomic data

From the Roche 2019 Annual Report

Linking genomic profiles with real-world data to attain deeper knowledge

Two newer developments are helping to advance the field of personalised medicine further. The first is comprehensive genomic profiling which can reveal as yet unknown genetic anomalies or mutational patterns driving a tumour's growth. The second involves the power of large quantities of curated medical data, such as therapy outcome data, captured from the electronic health records of numerous patients. The next frontier will be reached by integrating these two sets of data, namely genomic information and real-world clinical outcomes.

This is exactly what Foundation Medicine, focusing on the genetic profiling of cancer, and Flatiron Health, curating electronic medical records of cancer patients, decided to accomplish jointly, starting back in 2016. Having acquired these two companies in 2018, Roche strongly supports this effort. The continuously

- Prior example introduced a structural and technical solution without needing to change the culture of clinical care organizations
- Developed a suit of products to support an existing culture

Culture change

For organizations seeking to become more adaptive and innovative, culture change is often the most challenging part of the transformation. Innovation demands new behaviors from leaders and employees that are often antithetical to corporate cultures, which are historically focused on operational excellence and efficiency.

But culture change can't be achieved through top-down mandate. It lives in the collective hearts and habits of people and their shared perception of "how things are done around here." ....

Walker, B. and Soule, S.A., 2017. Changing company culture requires a movement, not a mandate. *Harvard Business Review*, pp.2-6.

How can we accomplish this culture change?

- Look to many of the entities that presented at this conference
- They have started towards this change

What change in clinical institutions might advance data science?

- Encouragement of embracing mathematical ideas and computers in young generations of healthcare providers or students considering healthcare professions
- A willingness to take risks and spend time developing data strategies
  - May need to hire a staff to start building infrastructure and interacting with clinicians before there is an immediate payoff
- Less territoriality
  - Great data analysis ideas may come from people all over an institution
  - Organizations must be truly open to allowing even those without a “leadership” title
- Alignment of payment incentives with innovation
  - Centers for Medicare and Medicaid Innovation gives opportunities for this
  - We really need good economic frameworks

- I have borrowed most of these ideas from Biotech companies

Questions / Comments