NAN
37TH ANNUAL CONFERENCE
OCTOBER 25-28, 2017
THE WESTIN BOSTON WATERFRONT
Neuropsychological Practice: A Clinician's Way Forward

Technology driven data collection: the role of biomedical informatics

Justin B. Miller, Ph.D., ABPP/CN
Cleveland Clinic Lou Ruvo Center for Brain Health
Las Vegas, Nevada

NAN
37TH ANNUAL CONFERENCE
OCTOBER 25-28, 2017
THE WESTIN BOSTON WATERFRONT
Neuropsychological Practice: A Clinician's Way Forward

Financial Disclosures

- I have no relevant financial relationships to disclose
  - Employee of: Cleveland Clinic
  - Consultant for: None
  - Stockholder in: None
  - Research support from: Avanir Pharmaceuticals, Keep Memory Alive
  - Honoraria from: National Academy of Neuropsychology

2

Goals of the presentation

1. Provide an overview of biomedical informatics (BMI) and application areas
   - *What exactly is “biomedical informatics?”*
2. Discuss the relevance of BMI applications and principles in the measurement of human behavior
   - *How is this useful for neuropsychologists?*
3. Present ideas for potential BMI applications that could be developed today to better leverage existing data sources
   - *How can we get started?*
4. Considerations for future development efforts
   - *Where do we go from here?*

3

What is “big data?”

“Big data is like teenage sex: everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it . . .”

- Dan Ariely, Ph.D., January 6, 2013
Professor of Psychology and Behavioral Economics
Duke University

5

Soooo . . . what is it?

Some of the earliest proposals have suggested a 3-dimensional framework[1]:

1. Volume: The actual depth and breadth of data available
2. Velocity: The rate of data flow, both into and out of the system
3. Variety: The diversity of data formats, types, and sources, etc.

Recent definitions have proposed an expansion to include:

4. Variability: The consistency (or lack thereof) of incoming data
5. Veracity: The quality of the data
6. Value: The relative value of the information and knowledge generated

[1]Laney, 2001

6

Is neuropsychology . . . big?

Current practices are not, even in the largest of clinics

- As technology based measurement become more prevalent, the influx of data will rapidly shift towards "big data" scale (higher volume, higher velocity, higher variety, etc.)

Neuropsychology is not the first field to grapple with an exponential increase in data (e.g., radiology, genetics)

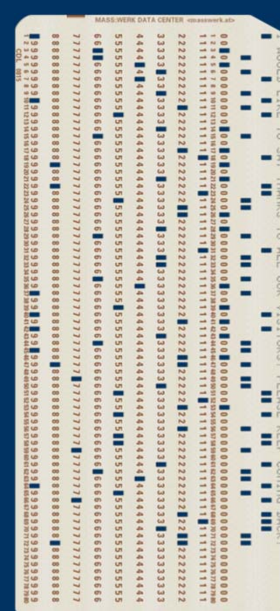
- Careful prospective planning about how data from these new sources will be collected, organized, and accessed will be necessary in order to derive new knowledge
- An inadequate digital data infrastructure to support the influx of data would be a significant barrier and impediment to translational research[1,2]

[1]Embi & Payne, 2009; [2]Payne, Johnson, Starren, Tilson, & Dowdy, 2005

Enter: informatics

Informatics generally refers to "science of <u>processing</u> data for <u>storage</u> and <u>retrieval</u>[1]"

- Origins of the field coincide with increasing access to computers in universities[2]
- Accelerated by development of relatively simply programming language in the 1960s (FORTRAN)[3,4]
- The continued (and ongoing) evolution of computing power has necessitated the concurrent evolution (and expansion) of data and information sciences



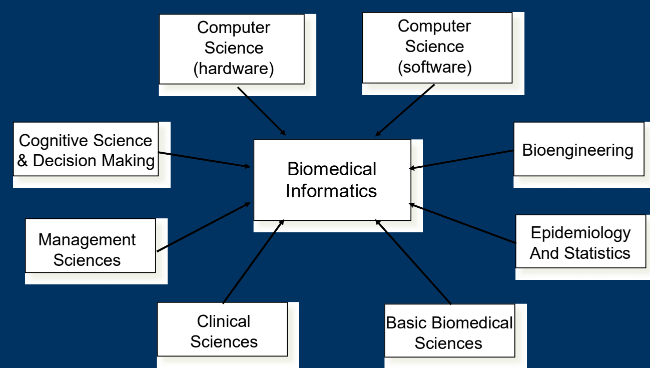
[1]Oxford English dictionary; [2]Anonymous, 1962; [3]Hagen, 2000; [4]Ledley, 1959

A bit of specificity: Biomedical informatics

Biomedical informatics (BMI) is an emerging independent discipline that is very broad and explicitly multidisciplinary

BMI draws upon[1,2]:

- Information technology
- Mathematics and statistics
- Computer science and engineering
- Biology
- Genetics
- Medicine



[1]Hagen, 2000; [2]Searls, 2010; www.amia.org

9

American Medical Informatics Association

- “The interdisciplinary field that studies and pursues the effective uses of biomedical data, information, and knowledge for scientific inquiry, problem solving and decision making, motivated by efforts to improve human health[1]”

10

A bit more specificity: biomedical informatics

BMI aims to utilize information technology to improve biomedical sciences

- Develop technology-driven resources to facilitate storage, access, use, and dissemination of health care data to generate new information and knowledge.
- Direct relationship between BMI and behavioral sciences as the information and knowledge is ultimately utilized by people to shape the healthcare landscape

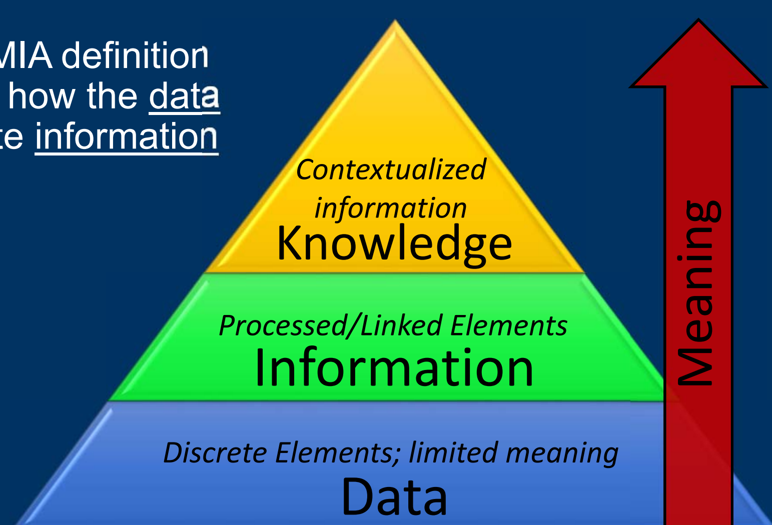
Biomedical informatics ≠ bioinformatics

[1]Kulikowski et al., 2012

Data vs. Information vs. Knowledge

A core component of the AMIA definition of biomedical informatics is how the <u>data</u> collected is used to generate <u>information</u> and derive new <u>knowledge</u>

- Hierarchically related
- Increasing degrees of connectedness
- Not synonymous



Data vs. Information vs. Knowledge

Data on their own are meaningless symbols or units (e.g. raw scores)

- A *data*base is nothing more than a simple collection of data with a very superficial organizational scheme and limited connectivity

Information reflects organized data with meaning attached to the individual unit (e.g., an age-corrected score)

- Most neuropsychology repositories are primarily data with some information

Knowledge is contextualized information, integrated with other known facts and larger meaning (e.g., cognitive domain labels)

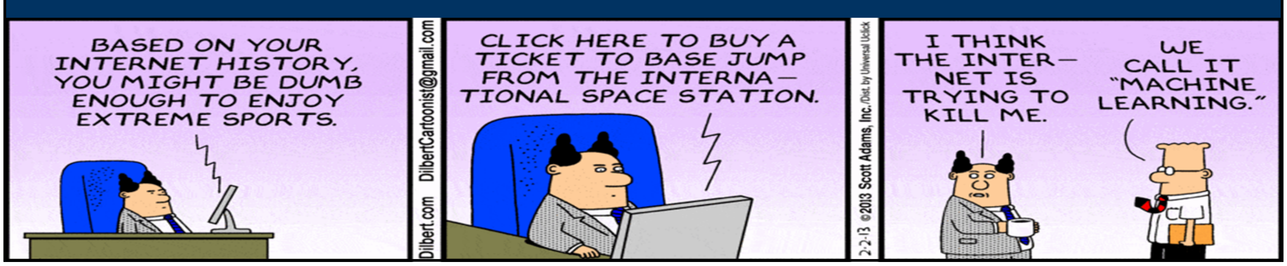
- New knowledge generation is driven by an active and intentional user effort (e.g., hypothesis testing)

13

BMI Application areas – where is this used?

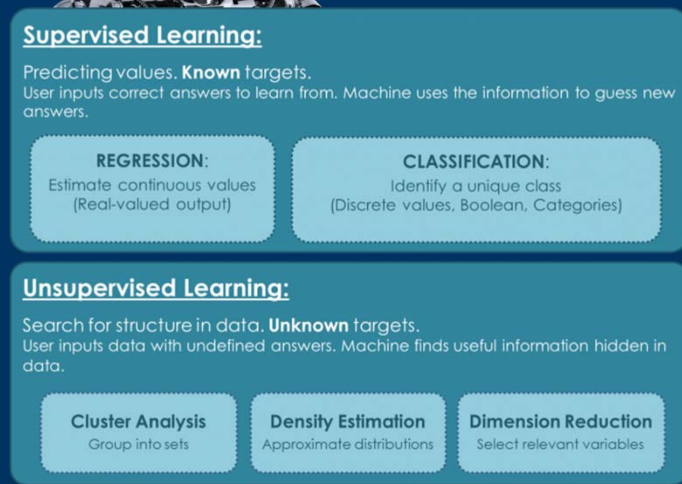
Translational bioinformatics: "The development of storage, analytic, and interpretive methods to optimize the transformation of increasingly voluminous biomedical data, and genomic data, into proactive, predictive, preventive, and participatory health"

- Example: data repositories that automate aggregation and facilitate integration of diverse data sources (e.g., biological data and clinical data)
- Bridging the gap between basic science and clinical knowledge



Machine Learning

- Teaching computers to learn without programming
- Provide set of data for training
- Develop algorithms based on supervised or unsupervised learning
- Test algorithms using new data
- Refine algorithms to improve accuracy
- Potential uses:
  - Differential diagnosis
  - Identification of biomarkers
  - Distinguishing subtypes of disease



15

BMI Application areas

Clinical informatics: “The application of informatics and information technology to deliver healthcare services. It is also referred to as applied clinical informatics and operational informatics.” (AMIA)

- How information is used in the delivery of healthcare services and providing information (e.g., clinical decision support tools; electronic medical records)
- Focuses on clinical applications

Clinical decision support (CDS): deliver filtered *information* to providers in order to enhance health and health care (HealthIT.Gov)

- Critically Appraised Topics, Clinical Care Pathways
- Passive data streams could alert care providers to interval changes

www.amia.org; www.healthit.gov/policy-researchers-implementers/clinical-decision-support-cds

16

BMI Application areas

Clinical research informatics: "The use of informatics in the discovery and management of new knowledge relating to health and disease."

- Typically relevant for clinical trials, but also relevant for continued utilization of amassed data and secondary use of clinical data for research (e.g., patient registries, collaborative knowledgebases)
- Counterpart to translational research informatics; focuses on clinical data vs. bench data
  - Distinction between CRI and TRI is getting very blurry

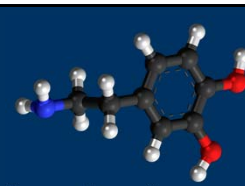
Definitions from www.amia.org

(critical) Importance of intended use

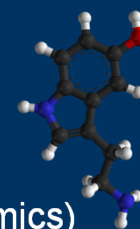
Each BMI application area delineates an intended use or purpose

- It is a central feature, designated at the outset of any application development
- Failure to do so will limit utility and may increases costs (e.g., redevelopment efforts; retrospective data entry)
- For example: Creating a database and populating it with data without thinking prospectively about how they will be used or by whom

Biomedical informatics - molecules



Bioinformatics: Historically used as the primary moniker for the field



- Now refers specifically to the study and application of BMI to molecular and cellular processes (e.g., genetics)
- BMI has driven the rapid advancement of precision medicine, primarily in determining drug treatments (i.e., pharmacogenomics)
  - Largely limited to oncology at this point (e.g., glioblastoma multiforme[1])
- Similar interventions may be possible for neurodegenerative disease or neuropsychiatric syndromes (e.g., UCLA Consortium for Neuropsychiatric Phenomics[2])

[1]Ene,& Holland, 2015; [2]www.phenomics.ucla.edu

Health informatics – individuals

Development of tools for study of person-level data

- Individual patient record with very simple functions (i.e., data aggregation, search/query, organization)
- Could be expanded via additional engineering to increase functionality
  - Track an individual patient's change over time (e.g., automated calculation of reliable change indices to track and monitor recovery or response to intervention)
  - Update treatment recommendations in real-time based on patterns of change

Public Health Informatics – populations

Focuses on population level data and societal health trends

- Aggregates data on a much larger scale to characterize health conditions at a societal level
  - E.g., automatically tracking primary presenting symptoms at emergency room admissions to monitor influenza outbreaks and alert hospital staff to implement increased precaution protocols
- Relies on large scale, systematic collection efforts that have only become feasible since collaborative cloud-based data collection has evolved

21

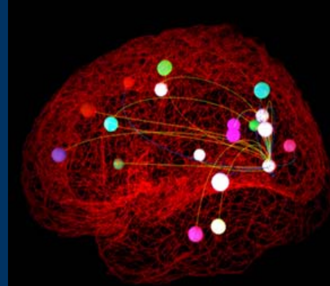
Neuroinformatics

Specialized BMI content area that focuses exclusively on central nervous system functioning

- Historically limited to imaging data
- Expanded to include broader diversity of data sources (e.g., functional data), molecular and cellular data, developmental data, cognitive data

Neuroinformatics is not entirely new to neuropsychology
(e.g., Jagaroo, 2009; Bilder, 2011; Parsons, 2016)

- <u>Cognitive Atlas:</u> establish the cognitive science ontology
- <u>Human Connectome Project:</u> understand neuroscience beyond local circuits (i.e., large scale neural circuits)
- <u>Disease-specific networks:</u> contain cognitive data but not central (e.g., ADNI, PPMI, TBIMS)



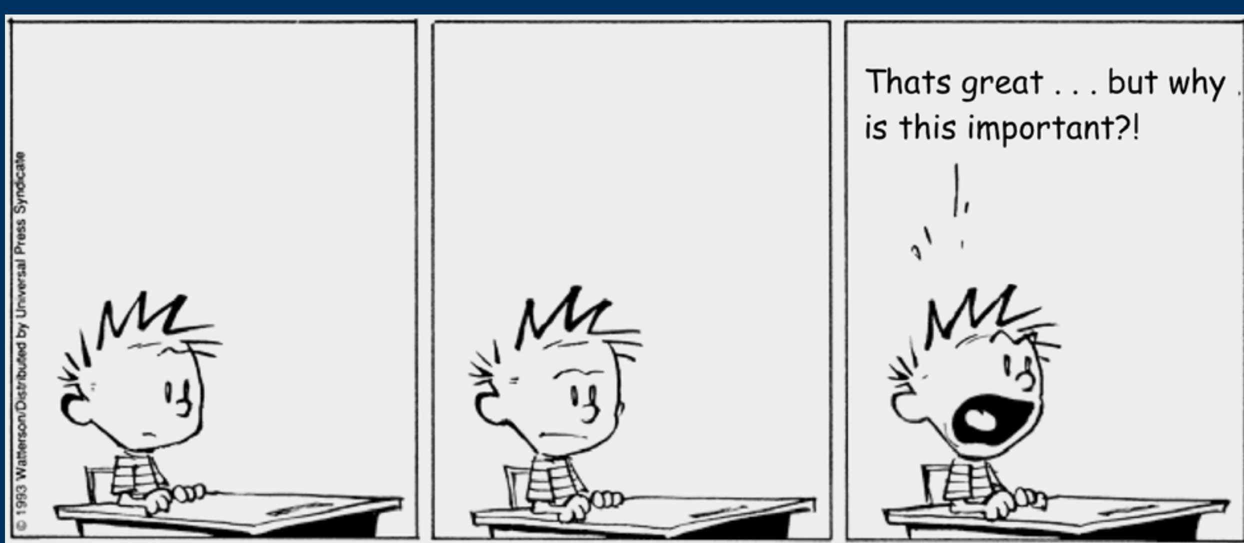
Neuroinformatics – basic applications

Many clinics and individual neuropsychologists likely (manually) maintain local databases

- But the connectedness with other health variables (e.g., blood labs) and other measures of CNS functioning (e.g., imaging biomarkers) is likely low
- Those that do incorporate other data sources likely rely on manual aggregation
- Multidisciplinary research datasets are purpose built for research studies

Cognitive and behavioral data can no longer be isolated from the rest of healthcare data!

23



24

Biomedical informatics has a lot to offer!

There is an entire field dedicated to the applied study of health information!

- It is as established as we are and is rapidly growing
- In the next decade, healthcare information technology job growth is expected to substantially outpace job growth compared to other industries[1]

Neuropsychology is already a highly data-driven field, despite using measures that are relatively data-poor (by modern standards)

- As quantification of human behavior becomes rooted in technology[2] there is critical need for an interdisciplinary perspective on data management

[1]Bureau of Labor Statistics; [2]Miller & Barr, 2017

What exactly *are* BMI applications?

Computer programs designed to increase usability of data to generate novel information and derive new knowledge:

- Electronic medical records (EMR)
- Clinical Decision Supports (CDS)
- Clinical Trial Management System (CTMS)
- Natural Language Processing (NLP)

Applications can vary in both scope and scale

- Local applications can simply aggregate and manage data from a single-provider or local user group
- Field-wide initiatives extend beyond individual users are may be multidisciplinary

Local Applications

At the local level (i.e., individual clinician/researcher or small workgroup):

- Improve patient outcomes and facilitate clinical decision-making
  - Pharmacogenomics – tailor medications based on genotype; already rapidly evolving
  - Pharmacophenomics – automatically titrating medications based on clinical symptoms
  - Automated Critical Appraisal (i.e., aCATs) – Integrate patient characteristics with published research to derive patient-specific risk/benefit profiles for treatments
- Increase clinical efficiencies
  - Automatically integrate information from EMR and data summaries directly into clinical reports
  - Create pipelines for automatic routing of referrals based on care pathways
  - Automate identification/screening of eligible research participants (e.g., clinical trials)

Local Applications – Increase NP productivity

A simple and straight forward informatics application that the individual clinician can develop is a basic database of clinical data obtained over the course of an evaluation

- Minimal programming experience using common spreadsheet software (e.g., Excel)
- Facilitate aggregation and organization of data according to user-defined parameters

With consideration of additional functionality at the outset of development, the utility can be greatly increased

- Automatically generate patient summary sheets
- Automatically convert scores to common metric and apply uniform set of descriptors
- Generation of graphical summaries of test data for report integration

Local Applications – Increasing NP information

Automated reliable change indices to monitor change over time

- With high throughput screening (e.g., remote assessment), significant changes could be automatically flagged and alert treating providers
- Could be enriched with additional clinical data from EMR (e.g., medication changes)
- With sufficient observations, the accumulation of practice-based evidence could lead to new care standards

29

Local Applications – Increasing NP information

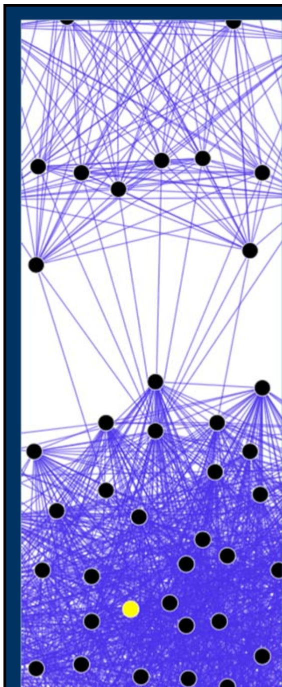
Diagnosis classification statistics could be derived for each patient

- Would initially require integration of regression formulas from published research
- Incorporating additional data sources (e.g., biomarker data), machine learning models could be developed to identify consistency of a cognitive profile with known pathology and refine prediction models

Measures of agreement (e.g., CCC) could be calculated against the existing standard within the battery to evaluate new tests

- Facilitate empirically-based test selection
- Incorporating biomarkers would guide empirically-based battery development

30



Collaborative databasing

Normative data

- "Living Norms" representing samples across the globe
- Test development and integrating novel measurement tools
- Base rate data relative to healthy controls and well-characterized clinical groups for both individual scores and cognitive profiles

Large scale analytics

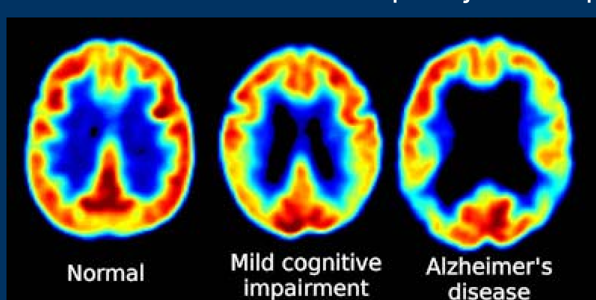
- Variable-centric methods (e.g., structural equation modeling): Empirically define cognitive constructs and their interrelationships both with other constructs and additional data (e.g., biomarkers)
- Person-centric methods (e.g., latent class analysis): multivariate taxometric analysis to identify distinct patient subgroups on the basis of a comprehensive panel of patient data (e.g., Clementz et al., 2016)

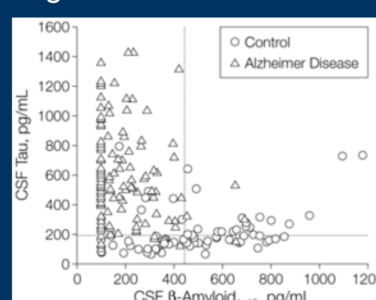
31

Collaborative databasing

Most neuropsychiatric disorders don't yet have a reliable biomarker[1]

- Diagnosis is based on clusters of observable symptoms, with pathology inferred
- Even though symptom clusters may not map well onto biological differences[2]
- Current medical records do a poor job of capturing mental health information[3]





4

[1]North and Suris (2017); [2]Clementz, et al. (2016); [3]Madden, Lakoma, Rusinak, Lu, and Soumerai (2016); [4]Sunderland et al., 2003

32

Automated Data Capture

Data reflecting cognitive functioning is readily available from everyday behaviors (i.e., “natural” data)

- Automated approaches to collecting predefined data points from typical, everyday behaviors need to be defined (i.e., capture it)
- This is conceptually distinct from “assessment,” which is an explicit, intentional, and *active* method by which behavioral data are obtained

Passive Data Collection (PDC) is an automated method of measuring and aggregating human behavior in a natural setting

- PDC methods are widely used in marketing (e.g., pop up ads, banners, etc.)
- Typically collected without our knowledge in an effort to preserve the “natural” interaction

33

Ecologically Valid Measures

PDC methods work exceptionally well when properly developed

Quantification of human behavior using PDC methods has the potential to dramatically shift neuropsychology

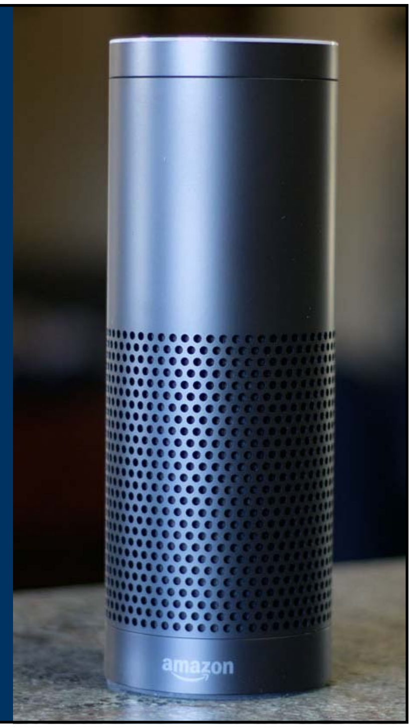
- These tools would generate data that are high velocity, high volume, and high variety; veracity (quality) and variability may be more dynamic
- The only way that meaningful information will be derived from these data is by development of BMI applications targeting these streams
- Integration with other non-cognitive/behavioral data sources may provide information that is more ecologically valid and a better representation of brain health

34

Novel Measurement Approaches

Smart devices (i.e., the Internet of Things)

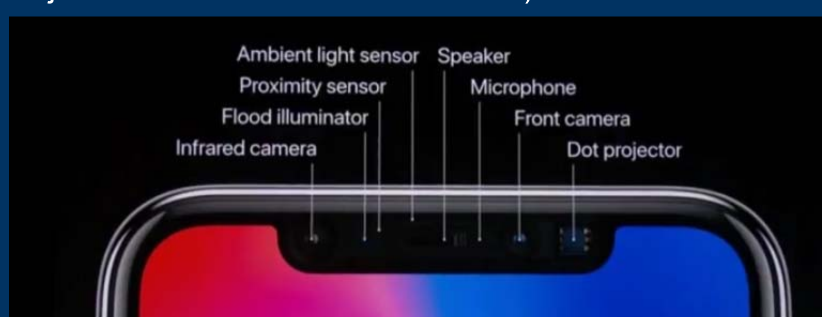
- In-home assistants (e.g, Amazon Echo, Google Home, Apple HomePod)
- Deploy rehabilitation interventions and monitor progress with integrated passive measurement
  - E.g., how often is a patient forgetting to turn off a light, or leave late for an appointment
- Update treatments in real-time based on passive data streams



Novel Measurement Approaches

Smart Phones

- Facial recognition sensors are becoming more widespread (emotion processing, eye tracking, facial asymmetry, etc.)
- Integrated Natural Language Processing (already done; we just don't have access to the data)

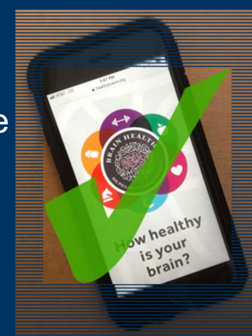


36

Personal Brain Record (PBR)

Current EHRs remain fragmented and ownership is retained by the healthcare providers

- Patient's rarely ever have complete access to their health records[1] and mental health records are especially sparse[2]
- The data that are contained reflect fractionated samplings of behavior that are disconnected from the remainder of health data
- Inconsistent nomenclature further limits utility



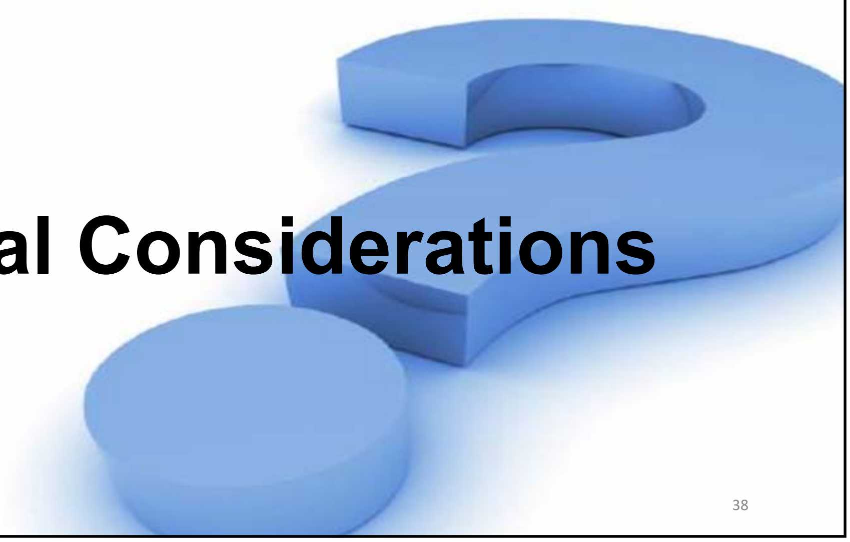
Pipeline for personal aggregation of mental health data, retained by the individual, that could be shared with healthcare providers

- Integrate with existing EHRs to increase depth of information
- Passive data collection methods (e.g., Instagram posts or Tweets)
- Remote clinical monitoring (e.g., medication adherence) and assessment (e.g., symptom inventories)

[1]Mikk, Sleeper, and Topol (2017); [2]Madden, Lakoma, Rusinak, Lu, and Soumerai (2016)

37

Additional Considerations



38

Data Security

Healthcare data breaches are common and getting more common (by mid-2017):

- 3.16M patient records breached
- 233 separate incidents
- Average of 325.6 days to discover the breach; 57 to report to DHHS
- 28% are inside jobs

Human error – not technology failure – still among the top causes!

- Title 21 CFR Part 11 – set of standards and controls for maintaining reliability, credibility, and equivalence of electronic records to paper records
- Compliance requires technological and procedural controls to ensure data are protected (e.g., audit trails, system validations/checks, electronic signature verifications)

40

Data Quality – Garbage in, Garbage out

Base rate data will be particularly useful (i.e., relative infrequency)

- Both in reference to a standardization sample and a clinical sample
- In addition to single-score prevalence rates, prevalence of pairwise scores could be determined (e.g., what is the base rate of these two scores in combination)
- Populating with well-characterized, and high-quality data from research trials may also help

- Development of quality assurance standards (e.g., file audits, double entry)
- Partially mitigated if novel tools input data automatically, reducing scoring errors, transcription errors, and entry errors
- Checks for validity also need to be established (e.g., identifying random responding)

41

Data Standards

The lack of field wide data standards is hugely problematic (not just a neuropsychology problem)

- Data are collected and stored in heterogeneous formants
- Data are described using a heterogeneous nomenclature

Data standards regarding how data are described, stored, and formatted need to be adopted (e.g., required by FDA for clinical trials)

- Standards only specify how to structure data for collection, storage, and analysis (e.g., variable naming systems, database metadata)
- Do not specify what to collect, what tools to use, or how to conduct assessments

42

Ethical considerations

The availability of data, as well as the tools to identify very subtle and previously undiscovered patterns of significance, have outpaced current ethical guidelines[1]

- What do you do when discovering a patient at risk for neurological or psychiatric syndromes?
- How deep does one dig to answer a clinical question?
- To what extent should the wealth of public data (e.g., Facebook, Twitter, Snapchat) be used in the study of public health?
- And many other things . . . .

[1]Gibney, 2017

Current efforts and future directions

Currently developing a cloud-based data repository for aggregation of neuropsychological data collected in the course of routine clinical service

- Built to improve clinical efficiency using an adaptive programming platform
- Integrate additional sources of non-cognitive data (e.g., brain imaging)
- Secondary purpose is to facilitate new knowledge discovery and data sharing
- Planned future developments to include:
  - Multi-disciplinary and multi-site collaboration
  - Platform for validating novel measurement tools
  - Automated data mining protocols


Thank you!
Questions, comments, suggestions: neuroinfo@neuropsychdata.com