QUANTORI

Computational Phenotyping and Analytics

Steven Labkoff, MD, FACP, FACMI, FAMIA
Global Head, Clinical and Healthcare Informatics, Quantori

Quantori

is an end-to-end data, technology, and digital solutions provider for the life science and healthcare industries.

We develop cutting-edge technology systems, applications, and infrastructures for biotech, pharmaceutical, and healthcare companies that accelerate drug discovery and improve patient outcomes.

Our industry domain and technical specialists lend expertise and support for ready-made tools as well as build complex high-level solutions from scratch.

What is Computational Phenotyping

QUANTORI

A new name for an old scientific need

A means to consistently identify groups of study or control subjects with a computationally-savvy method to generate a reproducible formula to identify groups

Also known as

- Groupers
- Inclusion/exclusion criteria

What is Computational Phenotyping

QUANTORI

Computational phenotyping refers to the use of computational techniques to identify and classify patient subgroups.

This can be focused on clinical and genetic data (though any kind of real-world data is in scope).

This approach allows for a more precise and personalized approach to healthcare by enabling the identification of distinct subtypes. This can be then applied to things like clinical care, clinical trials, or any kind of research where specific cohorts and controls are required.

What are Some Examples of Computational Phenotyping?

QUANTORI

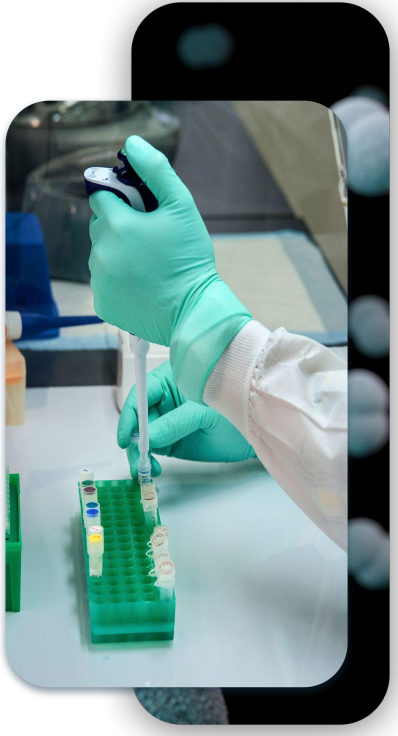
Creation of “like cohorts” for research (think clinical trial inclusion/exclusion)

- Study cases
- Control groups

Inclusion/exclusion criteria for a clinical study

Example:

- Show me all the patients with chronic kidney disease who were exposed to gentamycin for an infection in the previous 36 months, who also have type 2 diabetes and coronary artery disease



What are Some Examples of Computational Phenotyping?

QUANTORI

Alzheimer's disease:

- Computational phenotyping techniques can be used to identify distinct subtypes of Alzheimer's disease based on differences in clinical symptoms, biomarkers, and genetics.
- This can help tailor either study groups or even treatment plans and improve patient outcomes.

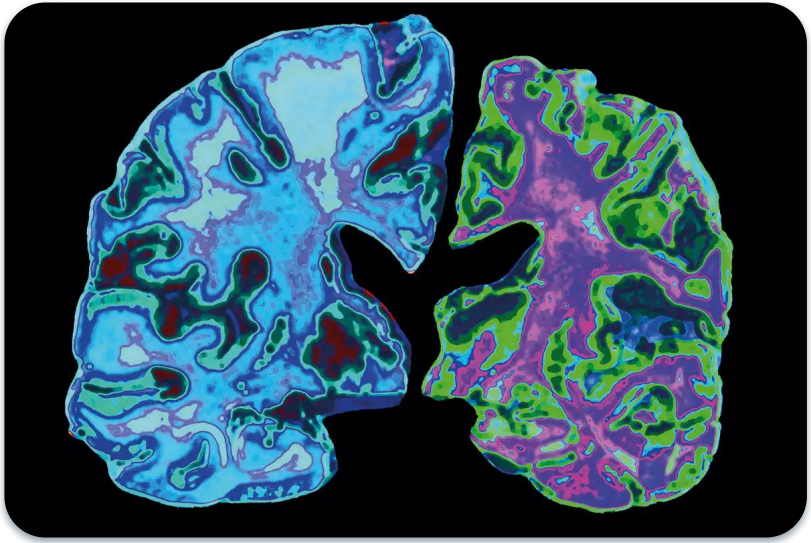


Image Reference: https://www.newscientist.com/article/2191814-we-may-finally-know-what-causes-alzheimers-and-how-to-stop-it/

Challenges in Computational Phenotyping

QUANTORI

- Data Quality
- Generalizability
- Privacy and Security
- Interpretability
- Standardization

Data Quality

The accuracy and completeness of the data used in computational phenotyping is critical to its success. If the data is incomplete, inaccurate, or biased, it can lead to erroneous phenotyping.

- EHR data is notorious for having missing data – making its use challenging for computational phenotyping
- If a clinical intervention is in use, it could lead to poor results and potentially harmful treatment decisions.

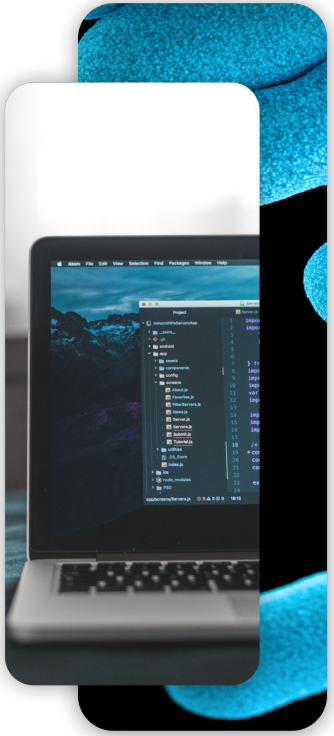
The Quality of cohort selection is predicated on the overall completeness and quality of the data sources. Poor quality data can lead to erroneous results

Privacy and Confidentiality

The use of large amounts of patient data for computational phenotyping raises concerns about privacy and security. It is important to ensure that patient data is anonymized and stored securely to protect patient confidentiality.

- Anonymization of data used in this manner is critical.
- Adding too many data sets together using things like tokenization can potentially complicate privacy concerns
- Patient consent for use in aggregated, tokenized data sets is unclear

We are entering a new era where the combination of multimodal data for use in studies via tokenization can create new issues around privacy and confidentiality



Standardization of Phenotyping Techniques

QUANTORI

There is currently no standardized approach to computational phenotyping, which can make it difficult to compare results across studies and institutions. Efforts to standardize phenotyping methods and data collection are needed to improve consistency and reliability.

- Every group/PI or analyst uses their own methods for creating phenotypes
- Because of this, comparing studies can be challenging because cohorts are seldom identical in construction

One major gap in informatics today is that of a common phenotyping mark up language

Interoperability

QUANTORI

- A lack of a common query (mark up) language makes creating repeatable, computable phenotypes challenging – especially across data types
- Efforts to develop more interpretable phenotyping algorithms and tools are needed to increase their clinical utility
- A lack of common phenotype libraries make comparing studies more challenging that needs be

Advances are being made using US-CDI and other standards ensure better interoperability of data – easing the ability to perform consistent phenotyping

The Ecosystem: Various Kinds of Data: A Partial List

QUANTORI

Personal Data

- Family Hx
- History of current illness
- Actual Medication list
- Detailed info on the patient journey

EHR

- H&Ps
- SOAP Notes
- Radiology Reports
- Pathology Reports
- Chem, Heme, & other Labs
- Rx's prescribed

Genomics

- NGS Sequencing
- Genomic Panels
- Whole Exome Seq
- FISH

'omics Data

- RNA sequencing
- Single Cell RNA Seq
- CyTOF
- Other immunological profiling

PROs

- Quality of Life
- AEs
- Patient experiences
- Outcomes

Claims

- What was done
- What was paid for
- How much was paid
- Time between events
- Claim Codes

Pharmacy

- Medication Lists
- Fill Data
- Cost data
- Approvals

The Ecosystem: Various Kinds of Data: A Partial List

QUANTORI

Personal Data

- Family Hx
- History of current illness
- Actual Medication list
- Detailed info on the patient journey

EHR

- H&Ps
- SOAP Notes
- Radiology Reports
- Pathology Reports
- Chem, Heme, & other Labs
- Rx's prescribed

Genomics

- NGS Sequencing
- Genomic Panels
- Whole Exome Seq
- FISH

'omics Data

- RNA sequencing
- Single Cell RNA Seq
- CyTOF
- Other immunological profiling

PROs

- Quality of Life
- AEs
- Patient experiences
- Outcomes

Claims

- What was done
- What was paid for
- How much was paid
- Time between events
- Claim Codes

Pharmacy

- Medication Lists
- Fill Data
- Cost data
- Approvals

Examples of Phenotyping in Action

The inclusion or exclusion criteria of most clinical trials – either interventional or observational are complicated:

- **(NCT-ID: NCT03910439) Avelumab in Combination With Hypofractionated Radiotherapy in Patients With Relapsed Refractory Multiple Myeloma**
  - Eligibility:
    - Patients must have previously treated RRMM refractory to, ineligible for, or intolerant of available therapeutic regimens known to provide clinical benefit (e.g, immunomodulatory [IMiD], proteasome inhibitor, and anti-cluster of differentiation (CD)38 monoclonal antibody-based treatments).
    - Presence of greater than or equal to 1 extramedullary plasmacytoma and/or lytic lesion amenable to XRT
    - Age greater than or equal to 18 years
    - Adequate organ function, and without serious comorbidity or disease (e.g., autoimmune disease), that would preclude concurrent systemic treatment or radiotherapy.

Some Challenges with Computational Phenotyping

QUANTORI

The inclusion or exclusion criteria of most clinical trials – either interventional or observational are complicated:

- **(NCT-ID: NCT03910439) Avelumab in Combination With Hypofractionated Radiotherapy in Patients With Relapsed Refractory Multiple Myeloma**
  - Eligibility:
    - Patients must have previously treated RRMM refractory to, ineligible for, or intolerant of available therapeutic regimens known to provide clinical benefit (e.g, immunomodulatory [IMiD], proteasome inhibitor, and anti-cluster of differentiation (CD)38 monoclonal antibody-based treatments).
    - Presence of greater than or equal to 1 extramedullary plasmacytoma and/or lytic lesion amenable to XRT.
    - Age greater than or equal to 18 years.
    - Adequate organ function, and without serious comorbidity or disease (e.g., autoimmune disease), that would preclude concurrent systemic treatment or radiotherapy.

Some Challenges with Computational Phenotyping

QUANTORI

- Many inclusion and exclusion criteria in ClinicalTrials.Gov are hopelessly convoluted
- Use of non-specific criteria make using them exceptionally challenging from a computational perspective
- Data stored in EHRs are also not well aligned for this exercise
- Analysis of these data sets – for study, data visualization or other kinds of analysis require a great deal of human intervention.

More work is needed to ensure data sets are better organized, cleansed, missingness addressed and consistent use of ontologies and common metadata to ensure consistent use of computational phenotyping

Some Challenges with Computational Phenotyping

QUANTORI

- Temporality in data sets can be tricky
- How do you describe the temporality in queries easily when there are undocumented breaks - and different visits
- Standards can change over time

Representation of temporal events can create nuanced challenges in phenotypes

Some Challenges with Computational Phenotyping

QUANTORI

Suggestion	Comments
Use of Common Data Representational models	Use of common data models for RWD (EHR, Claims etc) can provide better, more consistent creation of phenotyping formulas. Examples are the use of • FHIR • OMOP CDM (see OHDSI.org)
Use of common data standards	• Use of metadata standards such as ICD10, SNOMED-CT, RxNORM, LOINC can provide harmonized approaches to data representation – and secondarily better results from phenotyping
Use of standard ontologies and vocabularies	• Many common data standards are also ontologies that provide for hierarchical organization. SNOMED-CT is an example – whereby you can narrow or broaden your search by going up or down the ontological tree for queries.
Establishment of a common markup language	• Having a common markup language for creating phenotypes that could be shared could dramatically impove things like observational trials.

Some Challenges with Computational Phenotyping

QUANTORI

- Data is accurately and truthfully entered into the EHR for the sake of patient care
- Data in the EHR is easy to find and use
- Data in the EHR contains the TRUTH about actual patient care
- It is easy to aggregate data in the EHR can be used to study diseases and find new cures and insights
- Data from multiple EHRs can easily be combined to better understand population health
- The data in the EHR was collected primarily by and for the clinicians that take care of the patients

Some Challenges with Computational Phenotyping

QUANTORI

- Data is accurately and truthfully entered into the EHR for the sake of patient care
- Data in the EHR is easy to find and use
- Data in the EHR contains the TRUTH about actual patient care
- It is easy to aggregate data in the EHR can be used to study diseases and find new cures and insights
- Data from multiple EHRs can easily be combined to better understand population health
- The data in the EHR was collected primarily by and for the clinicians that take care of the patients

Each of these assumptions have FLAWS in today's Healthcare Ecosystem

Challenge: Analytics Depends on the Quality and Accuracy of Cohort Selections

QUANTORI

- Inaccurate insights
- Limited Generalizability
- Bias and discrimination
- Decreased model performance
- Reproducibility issues
- Difficulty with results interpretation
- Increased complexity
- Loss of trust

Clean Data Sources and Validated Phenotypes can mitigate much of these challenges

Contact Information

QUANTORI

Steven Labkoff, MD

Global Head, Clinical and
Healthcare Informatics

steven.labkoff@quantori.com
+1 (917) 599-7742