

Unlock Real-World Data with Machine Learning

*The Professional Society for Health Economics and Outcomes Research
European Conference, Vienna, Austria*

November 6–9, 2022

Speakers



Eric Q. Wu
Ph.D.

Managing Principal
Analysis Group, Inc.
Boston, MA, USA



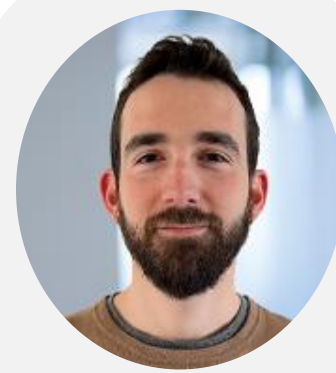
Jia Zhong
Sc.D.

Manager
Analysis Group, Inc.
Boston, MA, USA



Xiaochen Zhang
M.S.

Project Director
Beijing Huashu
Yihui Technology
Co. Beijing, China



Max Leroux
M.Sc.

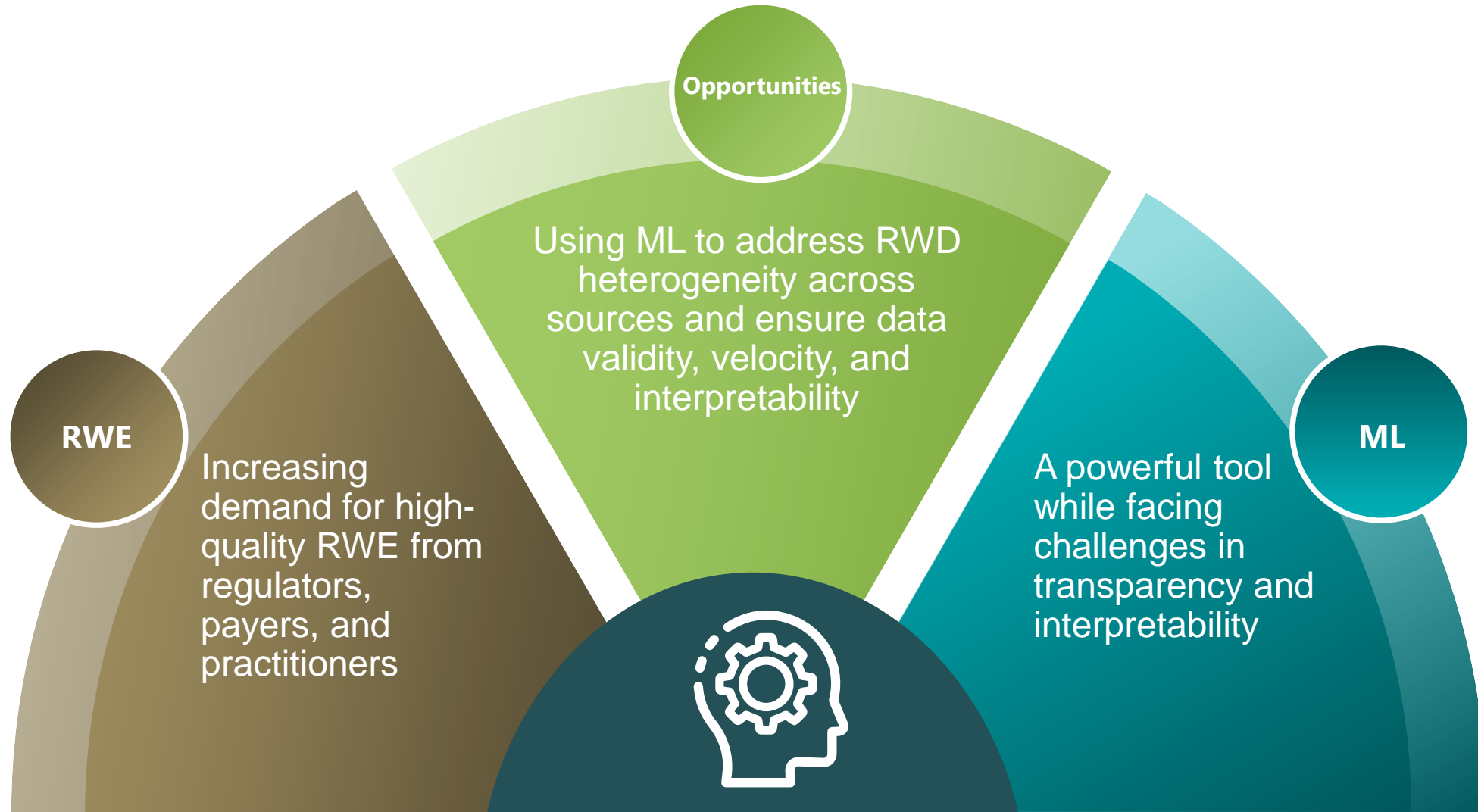
Director of Data Science
Analysis Group, Inc.
Montreal, Quebec,
Canada



Jimmy Royer
Ph.D.

Principal
Analysis Group, Inc.,
Montreal, Quebec,
Canada

Real-world evidence (RWE) and machine learning (ML)



Part I – Machine Learning in the Era of Big Data: Real-World Applications in Multicenter Studies in China

Jia Zhong, Sc.D.

Manager

Analysis Group

Email: jia.zhong@analysisgroup.com

November 6–9, 2022 Vienna, Austria

Agenda



Real-world evidence and multicenter study

01



Harnessing the power of real-world data (RWD) using ML tools

02



Examples of ML applications to empower multicenter RWD studies in China

03



Opportunities and future directions

04

Multicenter studies confer distinct advantages in real-world evidence

An increasing demand to support regulatory submissions and payer negotiation

Geographic coverage

Obtain understanding across regions and cities



Sample size

Maximize statistical power



Patient heterogeneity

Enable in-depth investigation of estimated effects in patient subgroups of interest

Recruitment

Facilitate recruitment by combining patient pools





Generalizability


Generate unbiased inferences beyond the study population

Challenges in real-world multicenter studies


ML technology provides efficient alternatives to empirical approaches


 Diverse infrastructural backgrounds
Different standards and process for de-identification

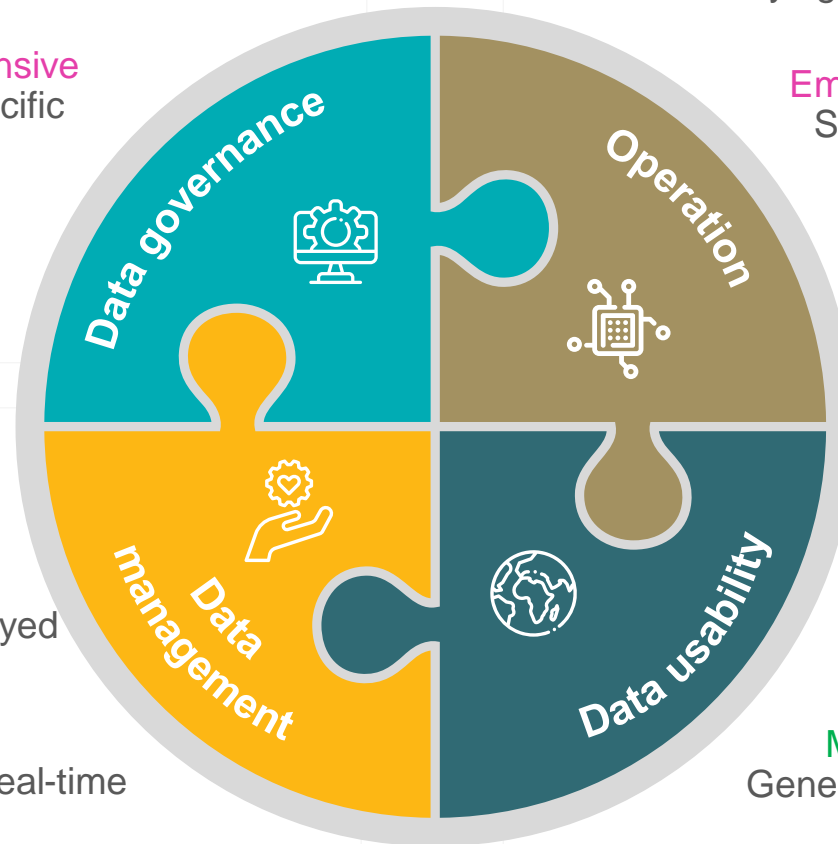
 **Empirical approach: Resource-intensive**
De novo data effort with center-specific setup/hardware


 **ML-powered solution: Flexible**
No requirement on existing environment


 Limited data linkage within center


 **Empirical approach: Unscalable**
Relying on manual review with delayed applications

 **ML-powered solution: Scalable**
Automated convergence enabling real-time analytics




Cross-department collaboration warranted
Varying operational and logistic requirements 

Empirical approach: Many speed limiters
Substantial training and on-site support 

ML-powered solution: Real-time
Standardized pipeline mechanizing instant data intake 

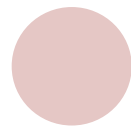
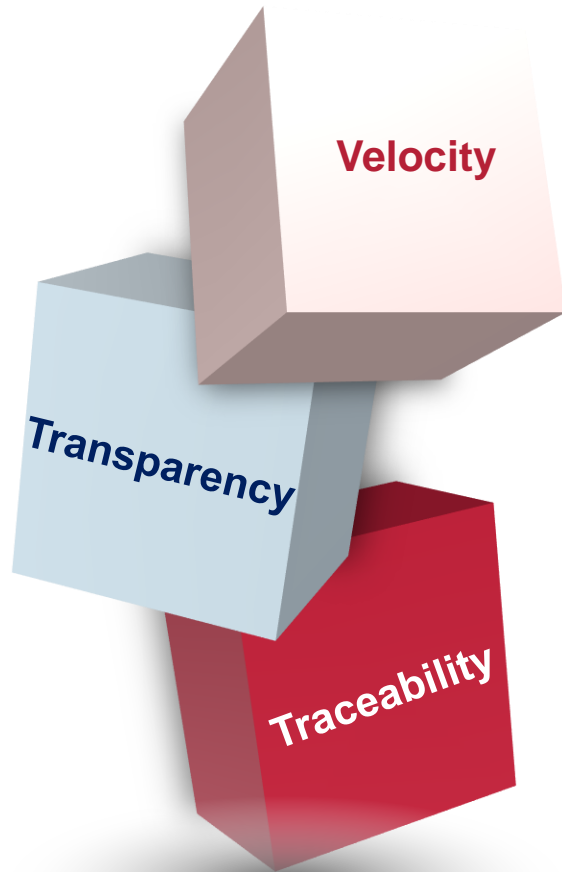
Inconsistent standards
Key variables not readily available 

Empirical approach: Limited value
Often inadequate for research 

ML-powered solution: Research-ready
Generalizable algorithms providing insights beyond existing metrics 

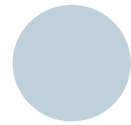
Machine learning enhances three key features of real-world data

Harnessing the power of RWD



Dynamically integrate various data sources

- Different data sources residing within the hospital systems represent different aspects of the patient journey
- An automated data capture system was developed to filter, process, and deposit tiny bits of data at a nearly constant rate



Uniform rules to establish consistency

- Combine consensus-based decision logic and data-driven optimization to create a set of transparent rules that can be used across centers
- Unify data standards and definition in multi-center settings



Maximize usability of historical information

- Provides a set of fact-based lenses to review data based on traceable data
- Can be adjusted as the clinical guidelines update and science advances

How does it work?

Data acquisition

Real-time data from hospital information systems involving multiple data sources and repositories



01

Data transformation

Transform unorganized raw data in various formats to structured and accessible data



02

Data cleaning

Resolve data conflicts, redundancy, and missingness



03

ML-powered dynamic disease model

A continuous iterating machine learning process leveraging mathematical modeling to tune and optimize model performance



04

Deployment

Deploy to the real-world data collection pipeline

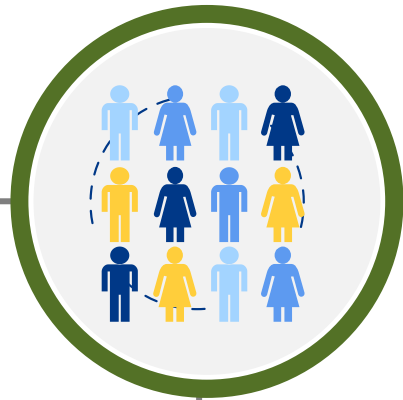


05

Example 1: National Longitudinal Cohort of Hematological Diseases in China (NICHE)

A flexible setup to enhance data velocity and mechanize information integration

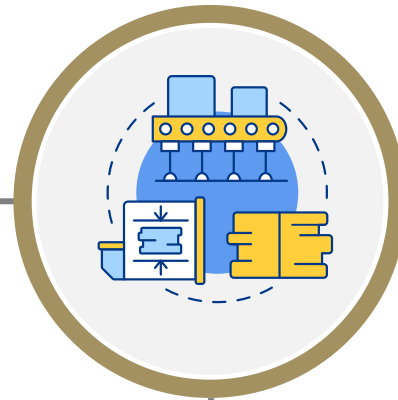
A prospective longitudinal multi-disease hematology cohort in China



26,906
enrolled



60,000+
with archived
biospecimens



11,808
in active follow up



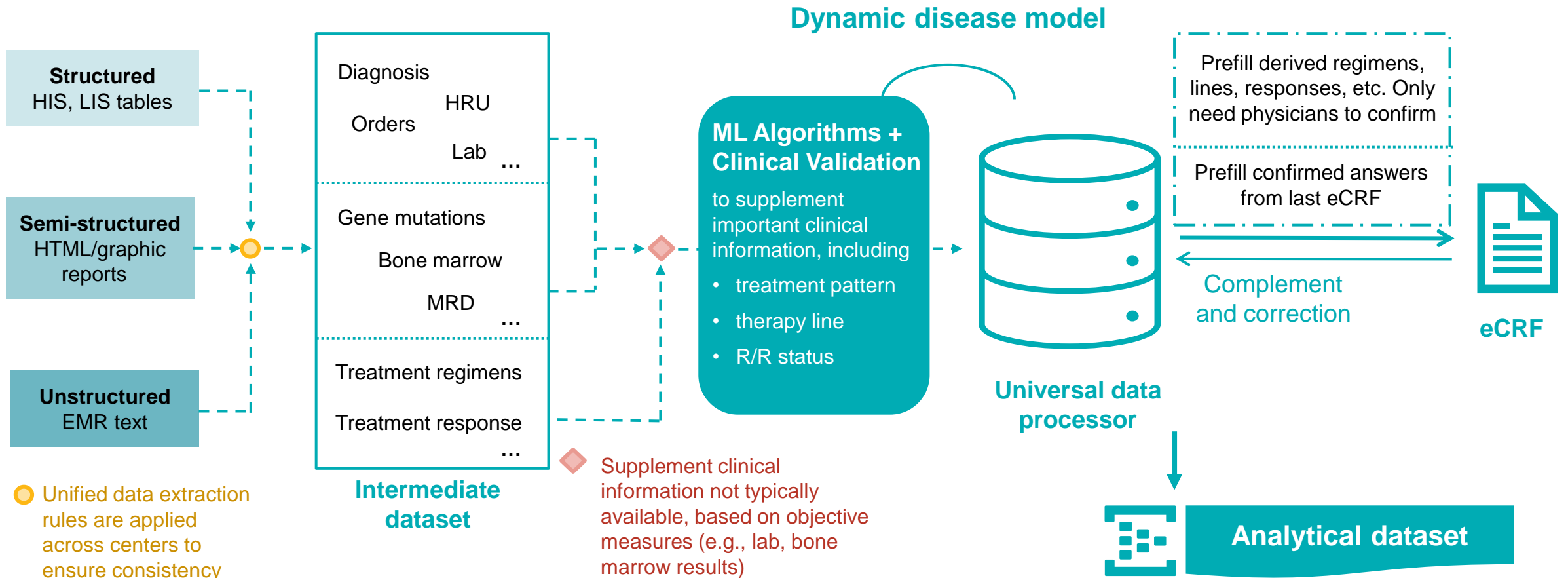
5.9 visits/year
727 days
average follow up



30+
hematological conditions

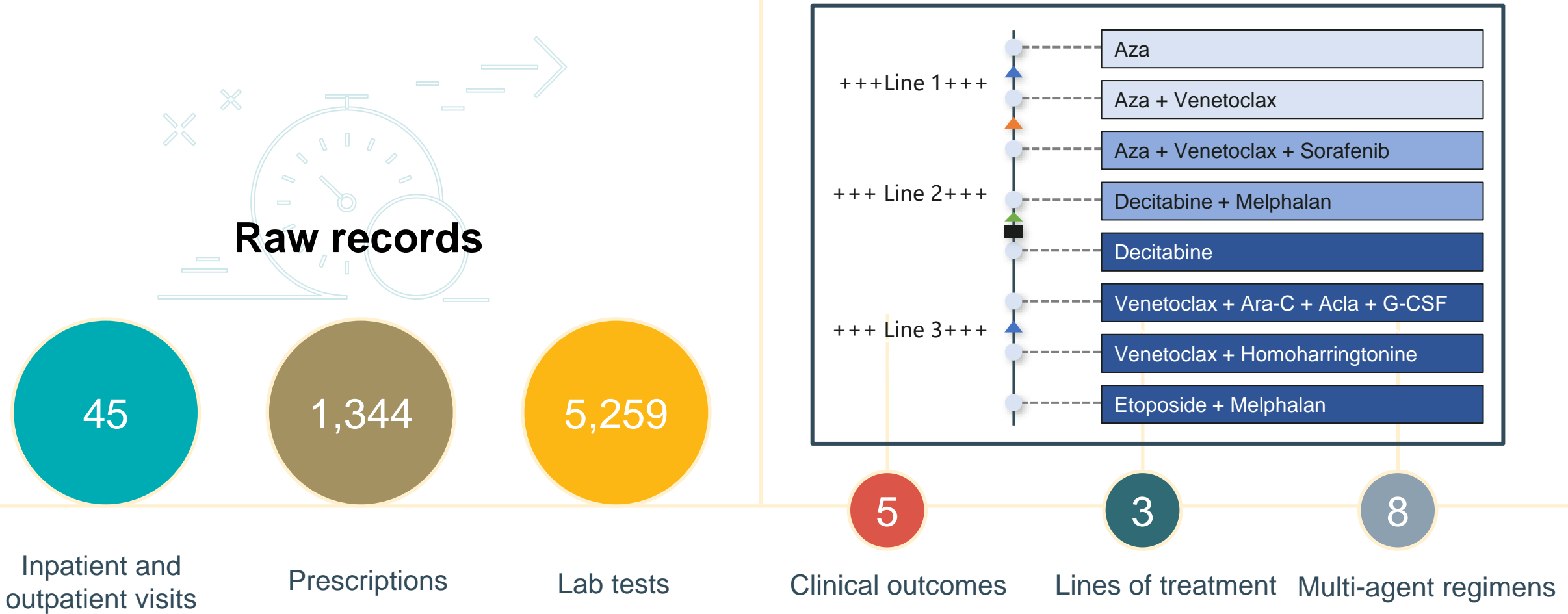
Data velocity: Integrate and streamline data firehoses

Capture, filter, and process high-volume RWD at a near-constant rate



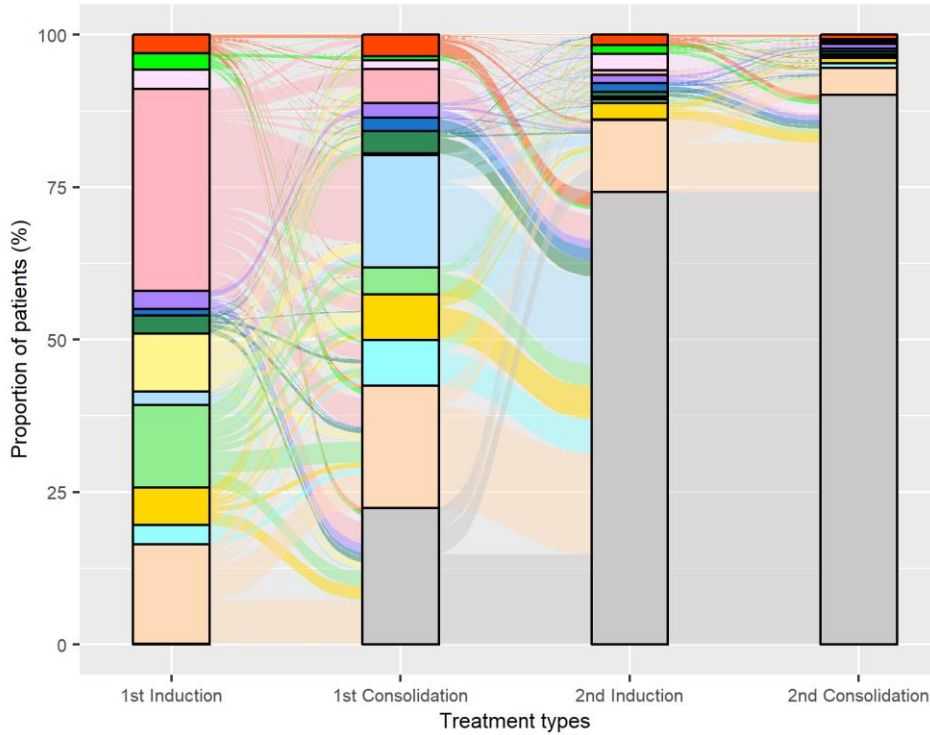
Transform historical raw data into a complete picture of the patient journey

Example: Mock data from one patient

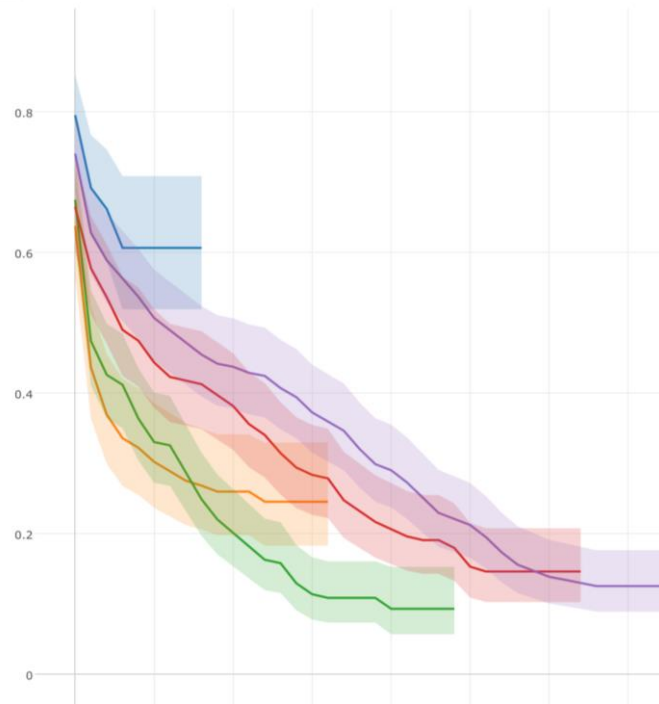


Generates research-grade dataset to support analytics

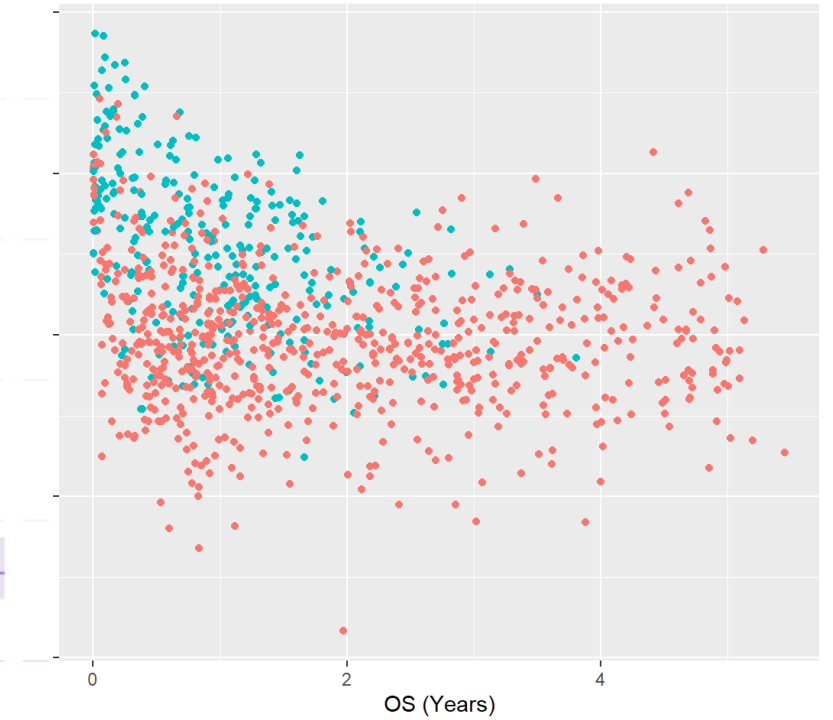
Sankey diagram of treatment sequences among patients with newly diagnosed AML



Treatment pattern analysis



Time to event analysis



Risk prediction

Example 2: Acute Myeloid Leukemia (AML) Real World EvidenCe (ARC) Initiative

Innovative solutions to unify data standards with transparency and traceability

AML Real world evidenCe (ARC) Initiative

A pioneer international longitudinal multicenter study engaging top-tier academic sites

US

10

MD Anderson Cancer Center

Dana-Farber Cancer Institute

Weill Cornell Medicine

Memorial Sloan Kettering Cancer Center™

GEORGETOWN UNIVERSITY

THE UNIVERSITY OF NORTH CAROLINA at CHAPEL HILL

THE UNIVERSITY OF ALABAMA AT BIRMINGHAM

MOFFITT CANCER CENTER

MEDICAL COLLEGE OF WISCONSIN

University of Colorado Cancer Center

Israel

4

RABIN MEDICAL CENTER
BEILINSON • HASHARON

RAMBAM Health Care Campus

HADASSAH UNIVERSITY MEDICAL CENTER
Founded by Hadassah, the Women's Zionist Organization of America

TEL-AVIV SOURASKY MEDICAL CENTER
MEDICAL EXCELLENCE AND COMPASSIONATE CARE

China

4

中国人民解放军总医院
PLAGH CHINESE PLA GENERAL HOSPITAL

中国医学科学院
血液病医院 血液学研究所

河南省肿瘤医院
郑州大学附属肿瘤医院

山东大学齐鲁医院
QILU HOSPITAL OF SHANDONG UNIVERSITY

Other countries

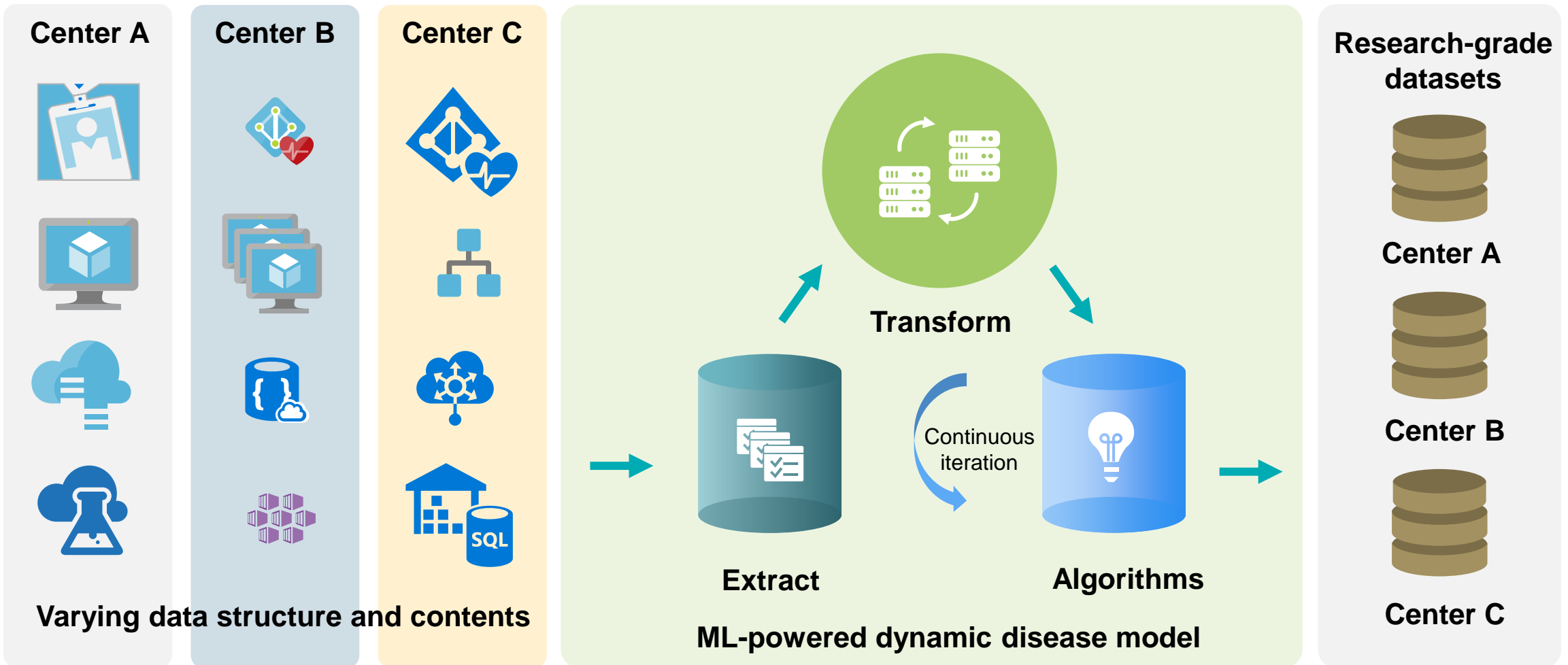
Expanding

Other European
Latin America
Japan
Other Asia-Pacific countries

...

ML-based data process pipeline to unify data standards and definitions

Improve the consistency, transparency, and traceability of RWD across hospitals in China



Data transparency: The “Necker cube illusion”

Inconsistent interpretation depending on focus and perspective

Clinical consensus

Healthcare data can have inconsistent or variable definitions, depending on practice. There may just not be a level of consensus.



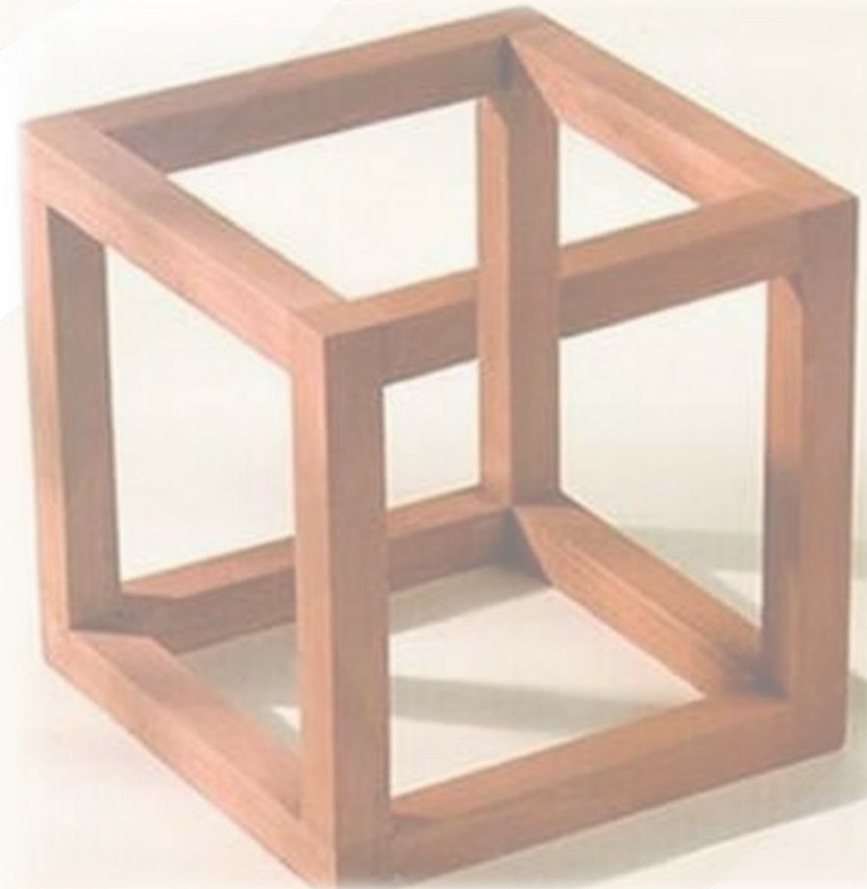
Emerging knowledge

Even when there is clinical guideline/gold standard, the consenting experts are constantly discovering new findings and knowledge.



Hitting a moving target

Continuous optimization of parameters in the data pipeline is key to create order out of chaos and constantly update as new understandings evolve.



How do you define refractory acute myeloid leukemia (AML)?

Will it simplify the issue when we have access to clinical facts?

Dr. A: "Not getting complete remission (CR) after induction chemotherapy."

Dr. B: "Well, it is if the disease worsens within 6 months."

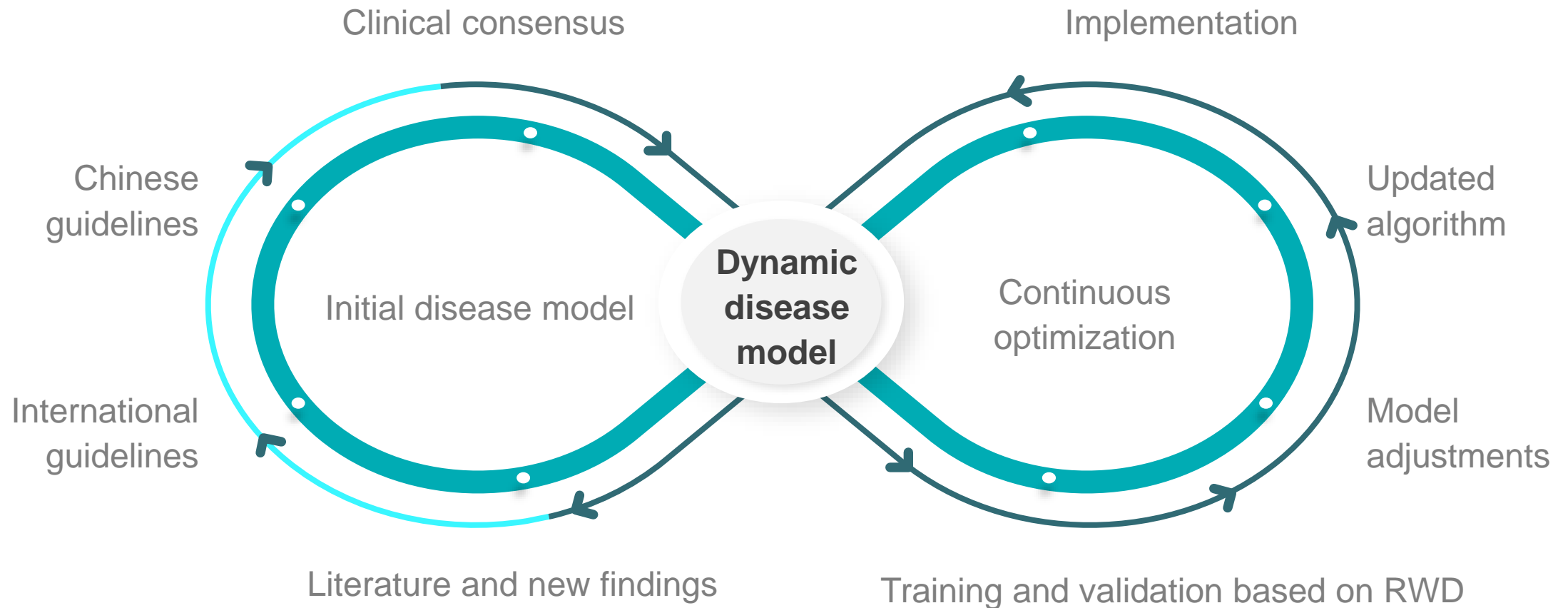
Dr. C: "No! Even with CR to induction, it is refractory if the disease worsens during consolidation therapy."

Dr. D: "Hmm, I think Dr. A is getting at it. But a patient is refractory if no CR after 2 cycles of induction."



Data traceability: The infinite learning loop in an evolving field

Enables systematic updates and adjustment to reflect new knowledge and consensus



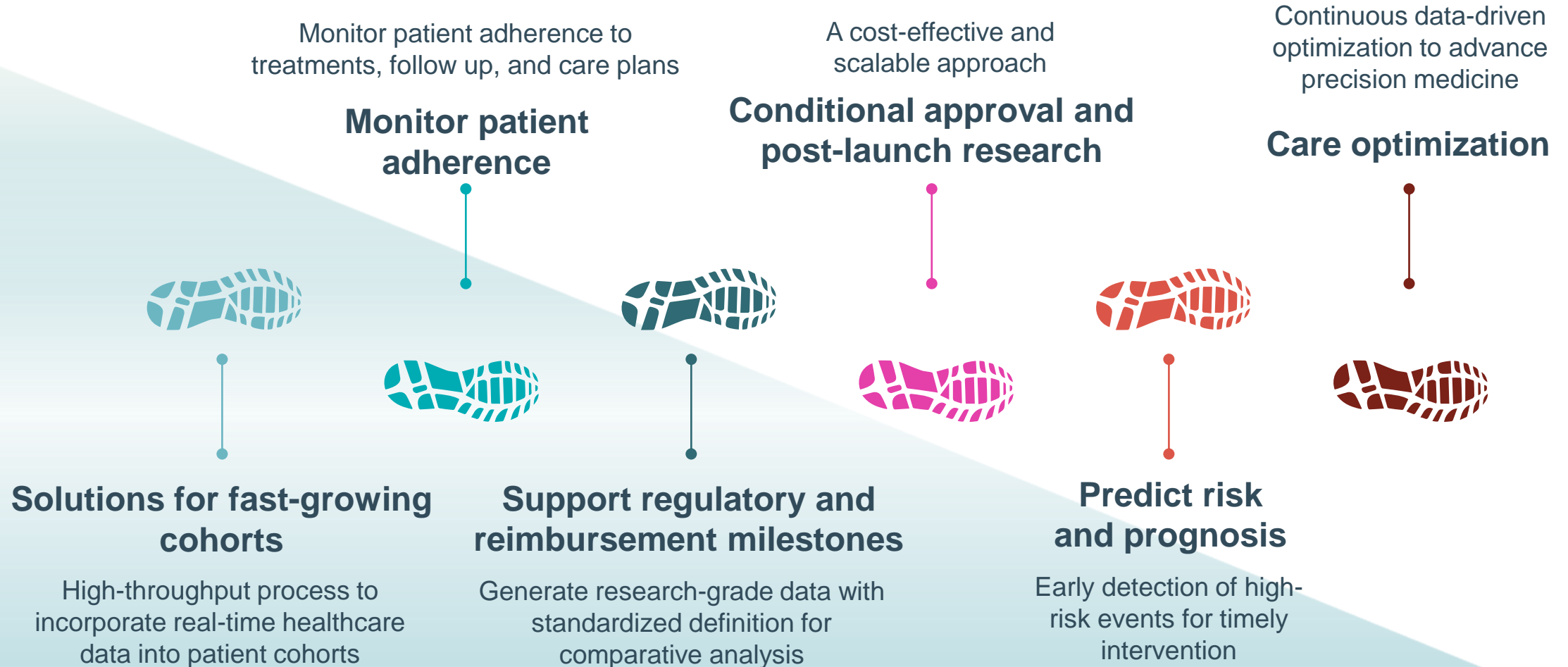
References:

National Longitudinal Cohort of Hematological Diseases, ClinicalTrials.gov, NCT04645199

Gong B, et al., Evaluating treatment patterns and outcomes for acute myeloid leukemia in adult patients in China - methodology considerations of RWE, Ninth Southern China Annual Congress on Pharmacoeconomics, 2021

Dynamic disease models: From data to insights

Opportunities and future directions of ML applications in RWE



Part II – Finding Order in the Chaos: Ensuring Data Relevance and Validity Using Machine Learning Tools

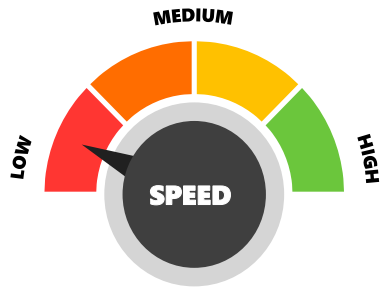
Xiaochen Zhang, M.S.

Project Director

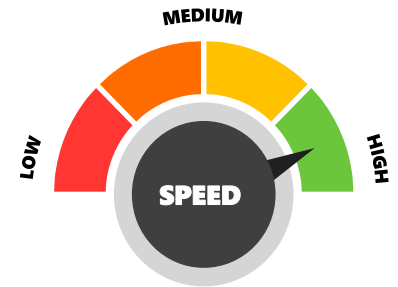
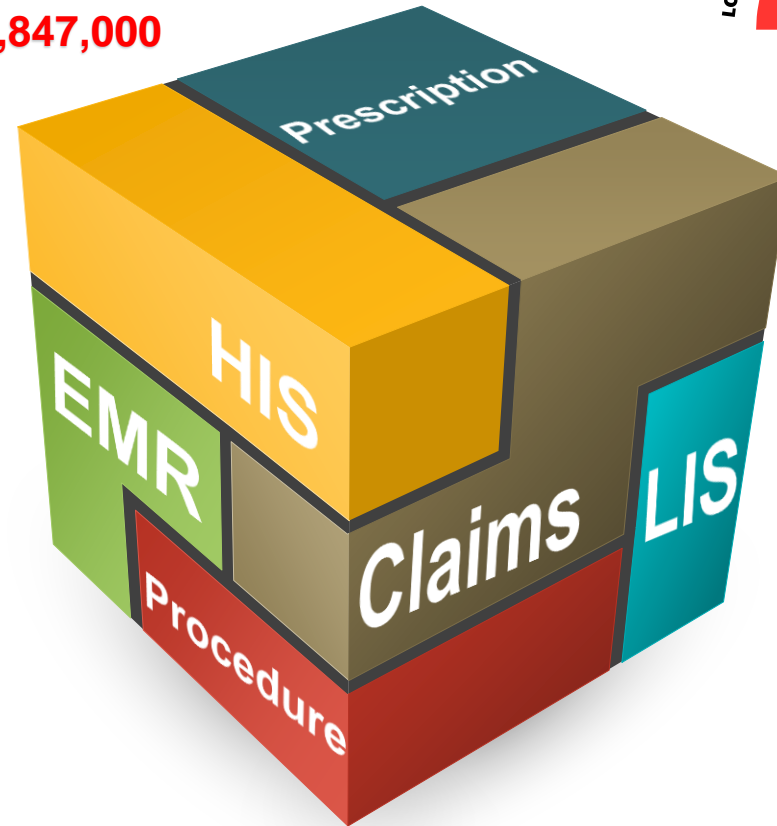
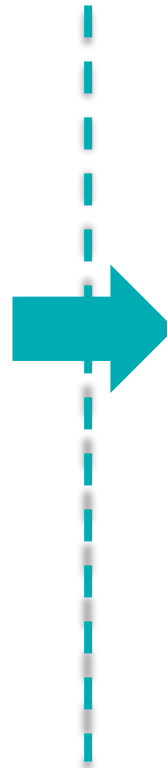
Beijing Huashu Yihui Technology Co.

Nov 6–9, 2022 Vienna, Austria

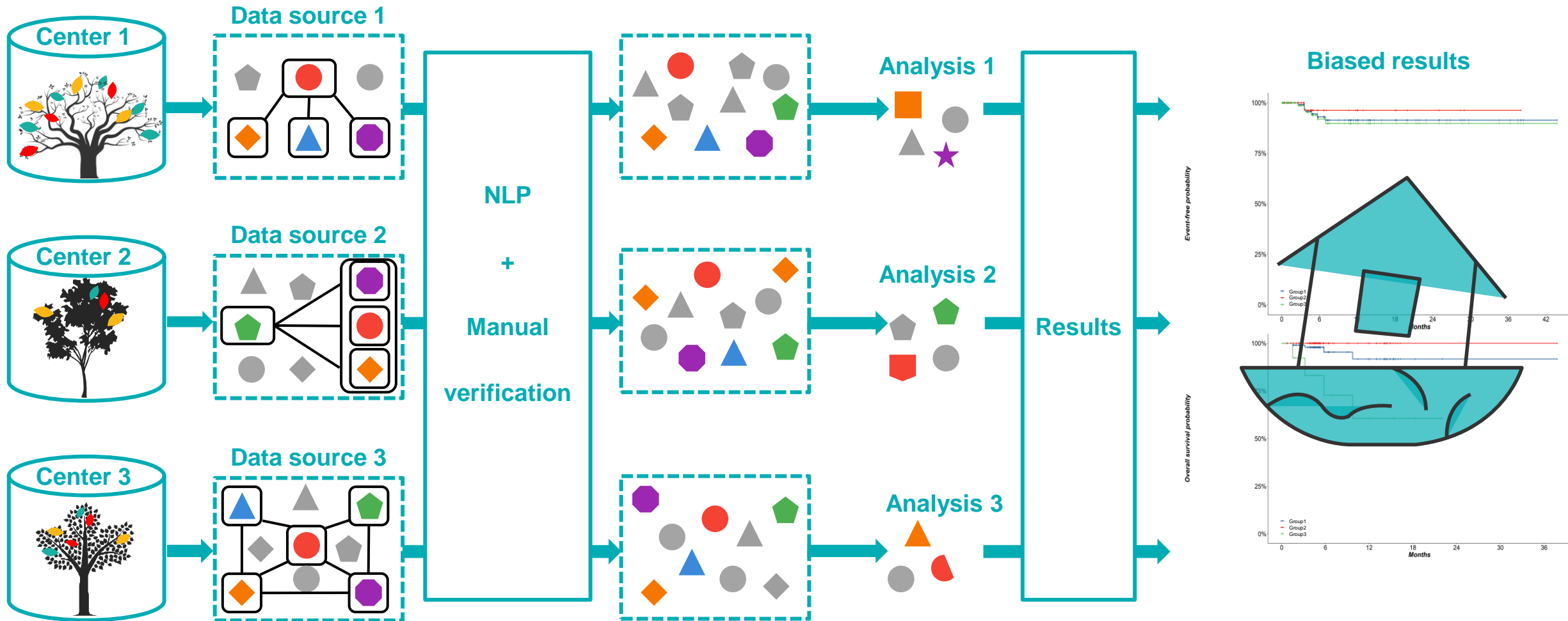
Huge amount of RWD



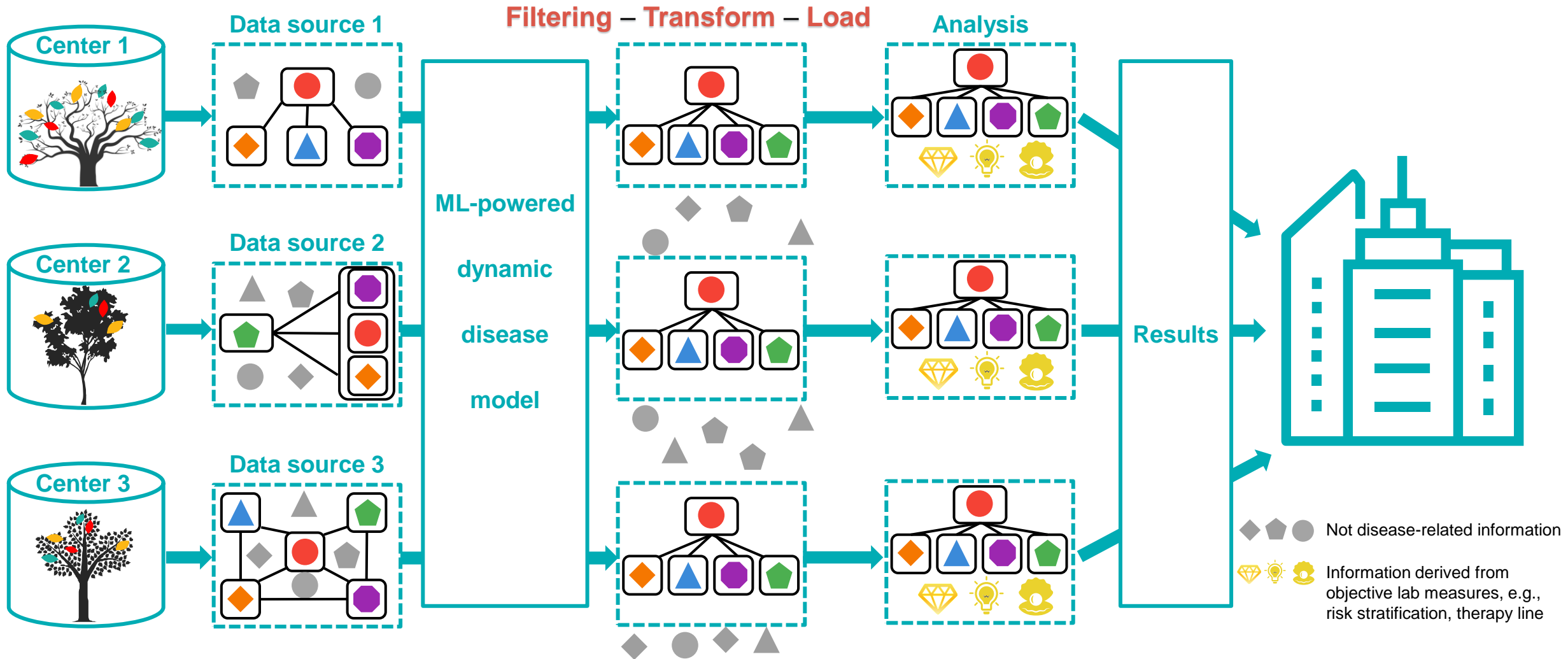
..... EMR 1,512,000 Prescription 1,785,000 Visits 270,000
 Procedure 455,000 LIS 1,194,000 Claims 17,847,000



Problems with RWD: Complex preprocessing for multiple centers

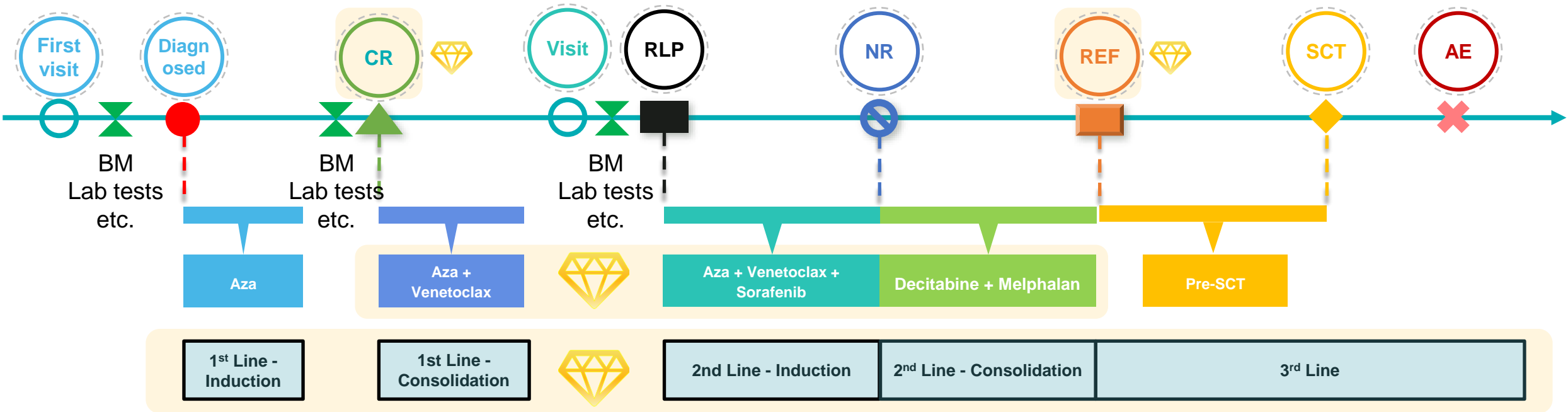



How do we ensure efficient integration of multi-source heterogeneous data?



What do we need? To transform raw records into a complete picture of the patient journey

Example: Mock data from one patient



 Valuable info from algorithm

Abbreviations: AE, adverse event; BM, bone marrow; CR, complete remission; NR, no remission; REF, refractory; RLP, relapse; SCT, stem cell transplantation.

References:

AML response criteria – CIBMTR forms instruction manual. <https://www.cibmtr.org/manuals/fim/1/en/topic/aml-response-criteria>

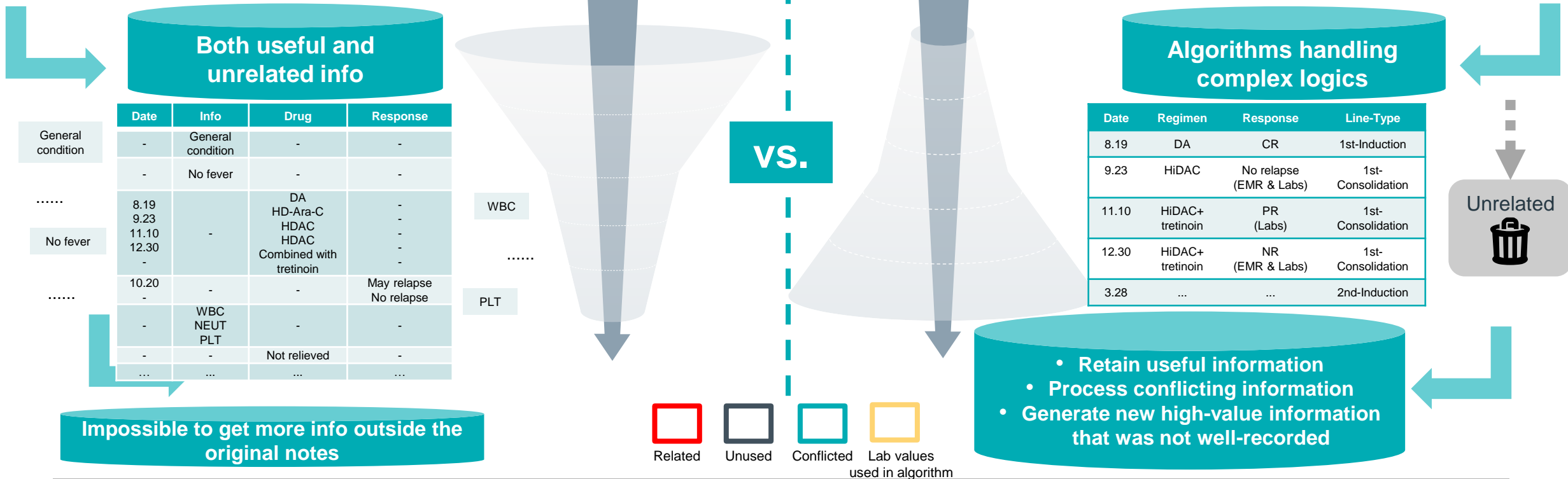
Ma, J. 中国复发难治性急性髓系白血病诊疗指南(2021年版). 中华血液学杂志2021年8月第42卷第8期 Chinese Journal of Hematology, August 2021, Vol. 42, No. 8

Benfa Gong, et.al. Evaluating treatment patterns and outcomes for acute myeloid leukemia in adult patients in China - methodology considerations of RWE, Ninth Southern China Annual Congress on Pharmacoeconomics

An empirical solution with challenges: NLP in healthcare research

Example: Mock physician chart from one patient

(3.28) The patient was admitted to the hospital for the first time in August 2021. The general condition of the patient was good, no fever and cough, good mental diet, and normal stool. Physical examination: anemic appearance,.....no swelling of both lower limbs. Gene mutation analysis showed that FLT3 ITD mutation, with a frequency of 35.3% Diagnosed acute myeloid leukemia M5b. 8.19, gave DA regimen for induction.... Quantitative detection of NPM1 gene mutation (type A) was 9.09%. 2021.9.23, HD-Ara-C consolidation therapy was given for 1 cycle; On November 10, 2021 and December 30, 2021, HDAC combined with tretinoin were given two times for consolidation chemotherapy. Lumbar puncture was performed for 3 times. 10.20 Bone puncture: 5% of the original cells, may relapse.. After relevant examination, judged that there is no relapse.. This time hospitalization: blood cell analysis (hospitalized venous blood): white blood cell WBC x 10 ^ 9/L ↓, absolute value of neutrophil - NEUT # x 10 ^ 9/L ↓.....showed that the patient was not relieved after last time treatment of acute myeloid leukemia. The patient... asked attention to 1. Go to the transplantation department for hematopoietic stem cell transplantation. The patient was a relapsed refractory leukemia with no remission; 2.3.24 Chest CT is normal, during pretreatment.....



Our approach: Knowledge dictionary + algorithm mode

Knowledge dictionary

- Cover multiple blood diseases
- Drug, procedure, lab tests, gene, etc.
- Continuous iteration

Name_cn	Name_en
阿糖胞苷	cytarabine
阿扎胞苷	azacitidine
高三尖杉酯碱	Homoharringtonine



Core algorithm

- Localized algorithm
- Solve scientific research problems, such as treatment line division
- Continuous iteration

2. 含中剂量 Ara-C 的诱导治疗方案: 高三尖杉酯碱 (HHT) $2 \text{ mg} \cdot \text{m}^{-2} \cdot \text{d}^{-1} \times 7 \text{ d}$, DNR $40 \text{ mg} \cdot \text{m}^{-2} \cdot \text{d}^{-1} \times 3 \text{ d}$, Ara-C 前 4 d 为 $100 \text{ mg} \cdot \text{m}^{-2} \cdot \text{d}^{-1}$, 第 5、6、7 天为 $1 \text{ g} \cdot \text{m}^{-2} \cdot 12 \text{ h}^{-1[9,14]}$ (证据等级 1a)。



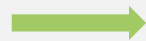
Data-driven findings

- Rely on a large amount of high-quality data
- Use ML to self-learn
- Instant feedbacks to the algorithm

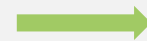
patient_id	regimen_cn	regimen_en
pd1	阿扎胞苷,阿糖胞苷,高三尖杉酯碱	HHT+Ara-C+AZA
pd2	阿扎胞苷,阿糖胞苷,高三尖杉酯碱	HHT+Ara-C+AZA
pd3	阿扎胞苷,阿糖胞苷,高三尖杉酯碱	HHT+Ara-C+AZA
pd4	阿扎胞苷,阿糖胞苷,高三尖杉酯碱	HHT+Ara-C+AZA
pd5	阿扎胞苷,阿糖胞苷,高三尖杉酯碱	HHT+Ara-C+AZA
...



HHT, AZA, Ara-C



HHT + Ara-C

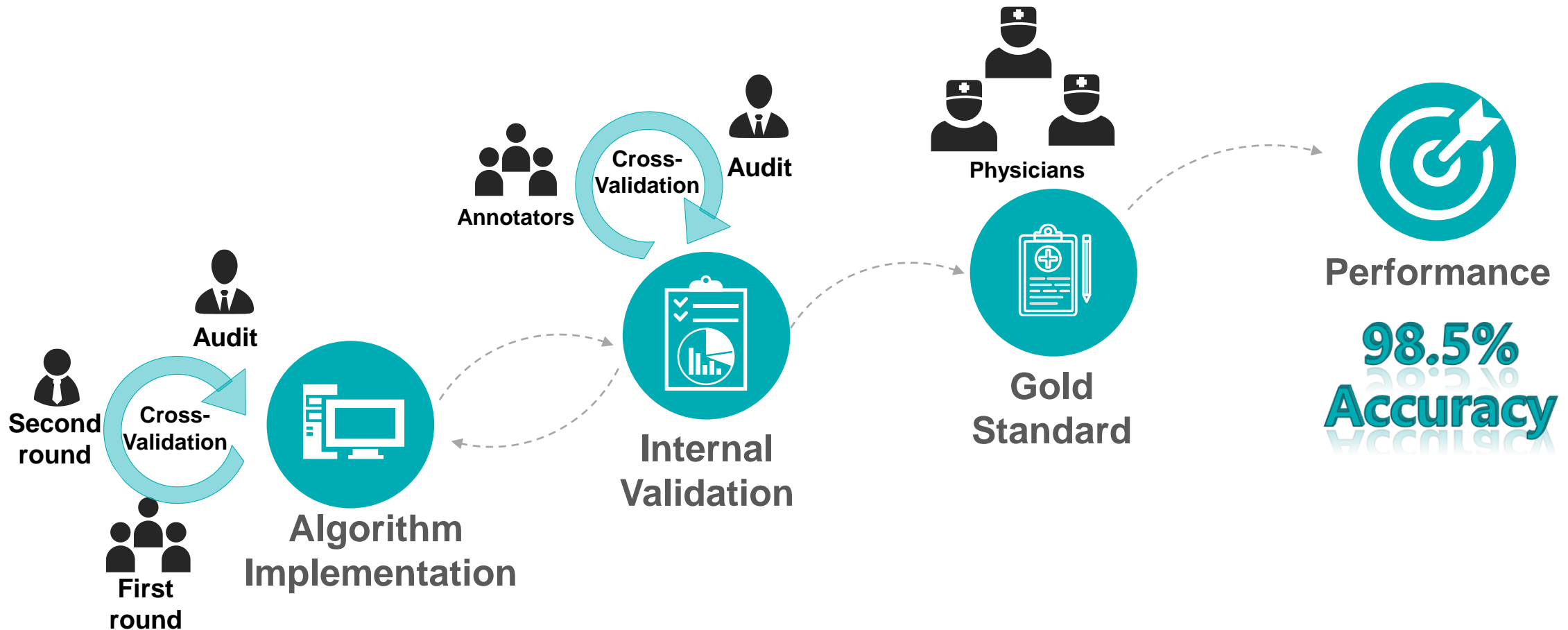


HHT + Ara-C + AZA



The outcomes: Algorithm performance

Example: Regimen algorithm of AML patients



What did we achieve? Algorithm outputs



- WHO classification
- FAB classification
- Risk stratification
- Treatment regimens in each treatment approach
- Treatment regimens in each treatment line
- Best response
- Relapse or primary refractory
- Transfusion independence
- Event-free survival

Our Algorithm	Others
✓	✗
✓	✗
✓	✗
✓	✗
✓	✗
✓	✗
✓	✗
✓	✗
✓	✗



Objective

- Focus more on objective values, such as lab results, etc.
- Better data governance



Reproducible

- Based on clinically relevant rules
- Better interpretability
- Better traceability



Flexible

- Remains flexible to accommodate varying parameters
- Flexible to add additional findings or tease out old markers

Our dynamic disease models are implemented in a variety of hematological conditions

Leukemia

32,000+

- Acute myeloid leukemia, AML
- Acute lymphoblastic leukemia, ALL
- Chronic myeloid leukemia, CML
- ...

40+

Lymphoma

9,300+

- Hodgkin Lymphoma, HL
- Non-Hodgkin Lymphoma, NHL
 - Diffuse Large B-cell Lymphoma, DLBCL
 - Mantle Cell Lymphoma, MCL
 - Marginal Zone Lymphoma, MZL
 - Follicular Lymphoma, FL
 - ...

Coagulation disorders

4,000+

- Hemophilia
- Von Willebrand disease, VWD
- Immune thrombocytopenia, ITP
- ...

Myeloproliferative disease

10,200+

- Myelodysplastic syndrome, MDS
- Myeloproliferative neoplasm, MPN
 - Polycythemia vera, PV
 - Primary myelofibrosis, PMF
 - Primary thrombocytosis, ET
 - ...

Myeloma

2,900+

- Multiple myeloma, MM
- Solitary plasma cell myeloma
- ...

Anemia

11,000+

- Aplastic anemia, AA
- Paroxysmal nocturnal hemoglobinuria, PNH
- Thalassemia
- Sickle cell anemia, SCD
- Iron deficiency anemia, IDA
- ...



Part III – Machine Learning and RWE in HEOR

Tools for transparency and interpretability

Max Leroux, M.Sc.
Director of Data Science
Analysis Group
Email: maxime.leroux@analysisgroup.com

Nov 6–9, 2022 Vienna, Austria

Problem Setting

- HEOR problems rely heavily on linear models (highly interpretable)
- However, to increase prediction accuracy, more complex models are needed – often at the cost of lower interpretability.

Objective

- Improve our understanding of complex models using state-of-the-art interpretability methods

Rationale for model interpretability in HEOR

GLM-based vs. machine learning-based models in HEOR

Traditionally used regression-based models in HEOR were transparent and interpretable

- Examples of models used in HEOR include GLM regression-based models:
 - linear regression for resources utilization
 - logistic regression for probabilities
 - Poisson regression for incidence rate

Interpretation through odds ratios, marginal effects

- These regression-based models all share the same structure : $E(Y|X) = \mu = g^{-1}(X\beta)$
- The results are usually transparent and interpretable:
 - marginal effects (b)
 - odds ratios (eb)
 - incidence rate ratio (eb)

Machine learning is more commonly used in HEOR

Along with RWE, machine learning tools are commonly used in HEOR

Recent publications using machine learning models in HEOR

- *Early Predictors of Sjögren's Syndrome: A Machine Learning Approach*, with J. Signorovitch, I. Pivneva, W. Huber and G. Capkun, *Value in Health* Vol. 22, Supp. 2 (2019)
- *Predicting clinical remission of chronic urticaria using random survival forests: machine learning applied*, with I. Pivneva, M-M. Balp, Y. Geissbühler, T. Severin, S. Smeets, J. Signorovitch, Y. Liang, T. Cornwall, J. Pan, A. Danyliv, S.J. McKenna, A. Marsland A and W. Soong, *Dermatology and Therapy* (forthcoming)
- *Development and evaluation of a predictive algorithm for unsatisfactory response among patients with pulmonary arterial hypertension using health insurance claims data*, with M. Gauthier-Loiselle, Y. Tsang, P. Lefebvre, P. Agron, K.B. Lynam, L. Bennett and S. Panjabi, *Current Medical Research and Opinion*, Vol. 38 (2022)
- *Development of a Multivariable Proxy Model for Six-Minute Walk Distance (6MWD) in Duchenne Muscular Dystrophy (DMD) Using Machine Learning Methods*, with N. Done, J. Iff, J. Signorovitch, D. Bertsimas, E. Henricson and G. McDonald, *Neurology*, Vol. 94 (2020)

Depending on the problem at hand and data used, different strategies have been used to make these models interpretable

- Parsimonious tree-based approach
- Most important predictors used in a logistic model
- Limited set of predictor used in a more narrowly defined model

Need for a unified framework

How can we make machine learning models more transparent and interpretable?

The solutions outlined work well

- In a specific context or within a specific problem set, simpler approaches can increase transparency and interpretability
- However, this case-by-case approach can be generalized through a unified approach

In the remainder of the presentation, we will focus on methodologies used to improve transparency and interpretability

Section 1: Interpretability methods

Section 2: Case study

Section 3: Final thoughts

Section 1: ML interpretability methods

Methods categorization

Interpretability methods can be thought of as either **model-specific** or **model-agnostic**.

- **Model-specific:** Interpretation methods that are *derived from the properties* of a model are said to be model-specific. The interpretation of *intrinsically interpretable models* like linear regression, logistic regression, GLM, and decision trees is always model-specific.
- **Model-agnostic:** Interpretation methods that can be *applied post-hoc, regardless of the structure of the underlying model* are said to be model-agnostic. They therefore work well with non-linear models (e.g., random forests, neural networks, etc.)

Interpretability methods can further be broken down into **global** or **local**.

- **Global interpretability** methods describe the distribution of the target variable based on the set of features. They inform on the marginal contribution of a feature across all possible coalitions (i.e., the entire model). Some examples include the importance of the feature or the directionality of the effect.
- **Local interpretability** methods describe each single instance's prediction individually. They inform on the relative effect of each feature in the particular prediction. This is especially useful for interpretability of inference.

Section 1: ML interpretability methods

Model-agnostic methods

Method	Global and/or local	Regression (e.g., RU)	Classification (e.g., probabilities, incidence)	Categorical covariates	Numerical covariates
Feature importance	global	✓	✓	✓	✓
Partial dependence plots (PDP)	global	✓	✓	✓	✓
Accumulated local effects	global	✓	✓	✓	✓
Global surrogate	global	✓	✓	✓	✓
Shapley values	global & local	✓	✓	✓	✓
Local interpretable model-agnostic explanations (LIME)	global & local	✓	✓	✓	✓

Section 2: Sample case study

Context

Objective

Build a prognostic model predicting the one-year disease progression of patients

Competing models performance

Model	Root mean squared error	Mean absolute percentage error	R ²
Lasso regression	4.144	3.249	0.221
Lasso regression (with interactions)	3.996	3.113	0.294
Random forest regression	3.910	3.097	0.307

Data

- **Patients:** ~ 800
- **Features:** > 20 Features
- **Outcome:** Disease (index) change over one year (real number)
- **Train-test ratio:** 4:1

Interpretability methods used

- Partial dependence plots (main and interaction effects)
- H-statistic
- Shapley values (global and local)

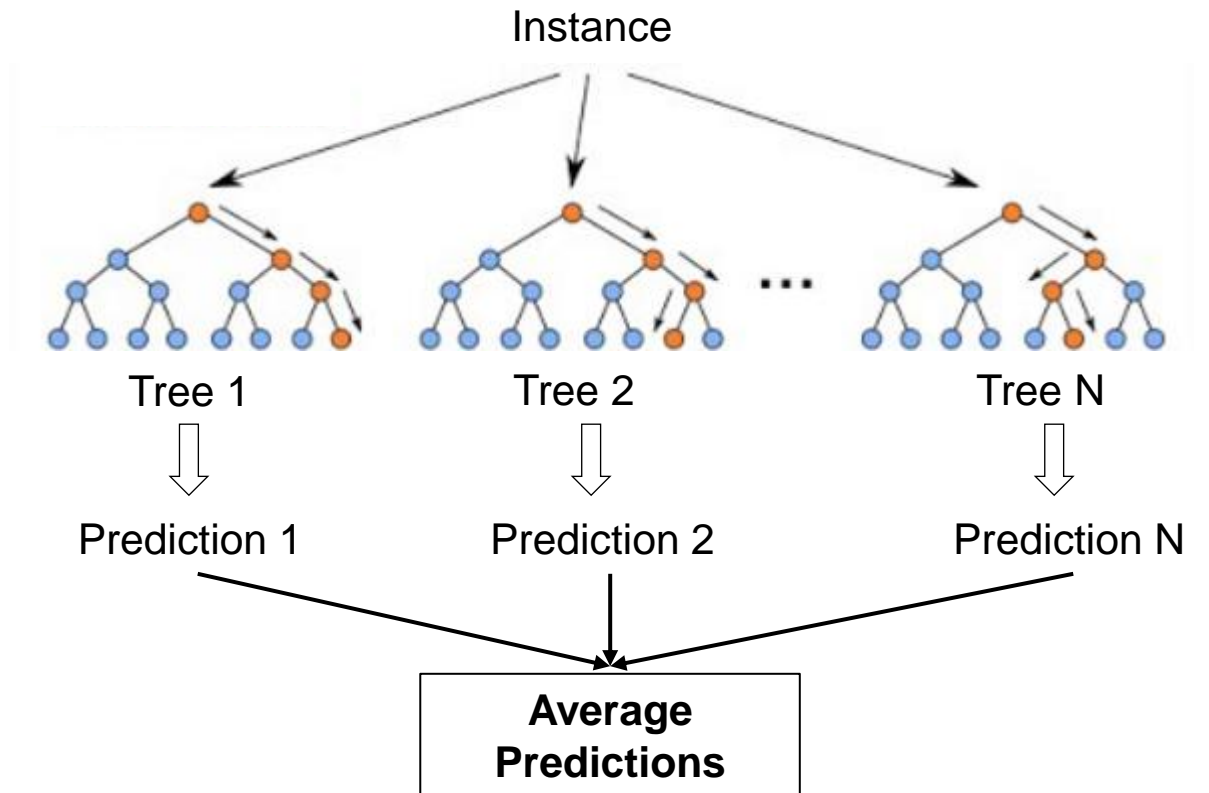
Section 2: Sample case study

Context

Lasso regression with interactions

Feature	Coefficient
Intercept	9.292
Age	-2.496
Feature2	7.196
Feature3	0.000
...	...
Feature8:Feature7	0.189
Feature8:Feature8	0.000

Random forest



Section 2: Sample case study

Feature importance

Feature importance shuffles each feature one at a time and measures the increase in model error as importance value of that feature.

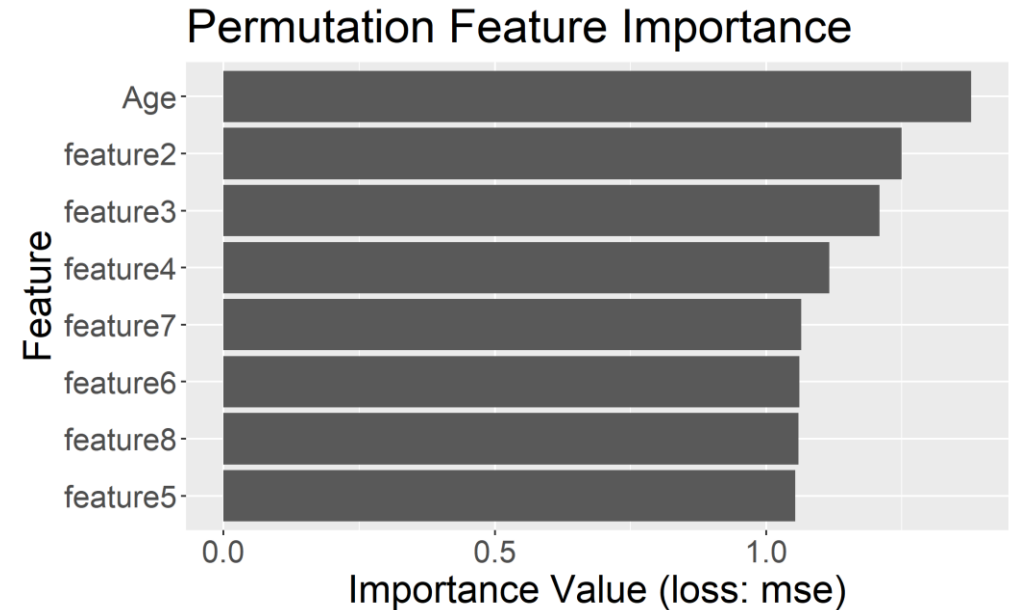
- **What it does:**
 - Measures importance value of each feature and rank accordingly
- **Comparable method:**
 - Average decrease in impurity
 - Importance value of a feature in the trained model is calculated by taking the average of the accumulation of impurity decrease within each tree in cases of multi-model architectures.

Advantage

- Easy to interpret
- Can be computed on a left-out test set

Limitations

- Correlated features can decrease the importance of the associated feature
- Need access to true outcome



Section 2: Sample case study

Partial dependence plot (PDP) – main effect

PDP reduces the complex model to a simple function that depends only on 1 or 2 features by **averaging** the effects of the other **features over the marginal distribution**

■ **What it does:**

- Estimates first order feature effects on prediction.
- Shows the dependence between the target and the feature(s), marginalizing over the values of all other features.

■ **Assumptions:**

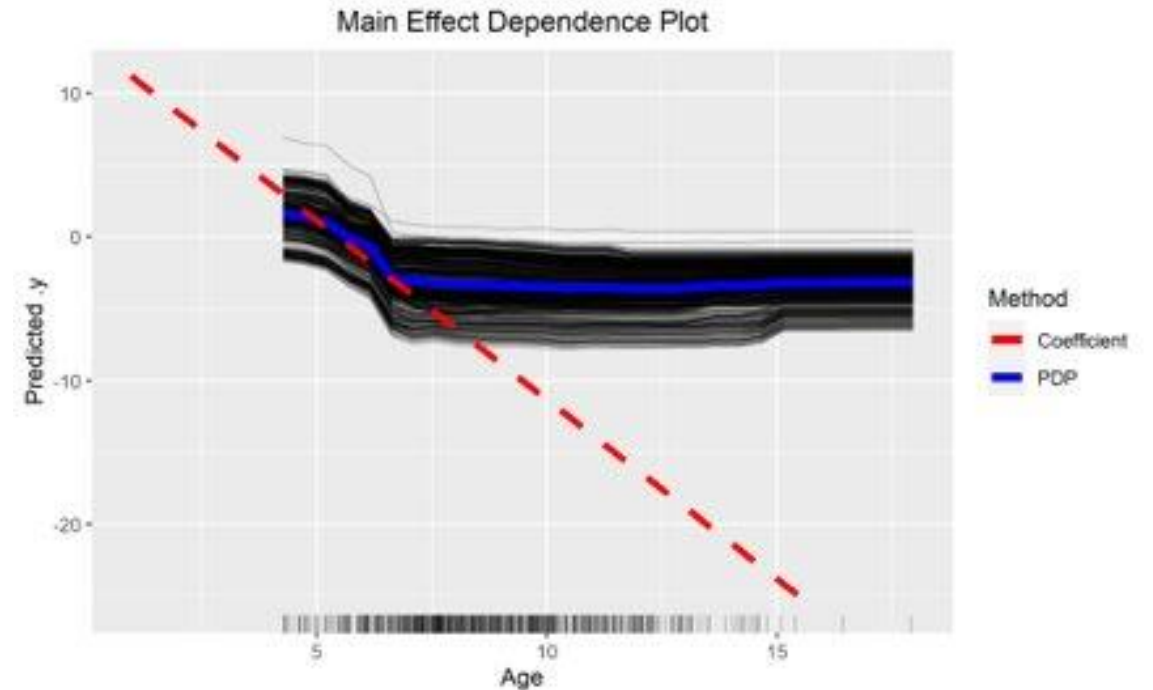
- Features are uncorrelated

Advantage

- Easy to implement and understand
- Can be paired with ICE plot

Limitations

- Stressing one feature potentially generates *unrealistic regions*



Section 2: Sample case study

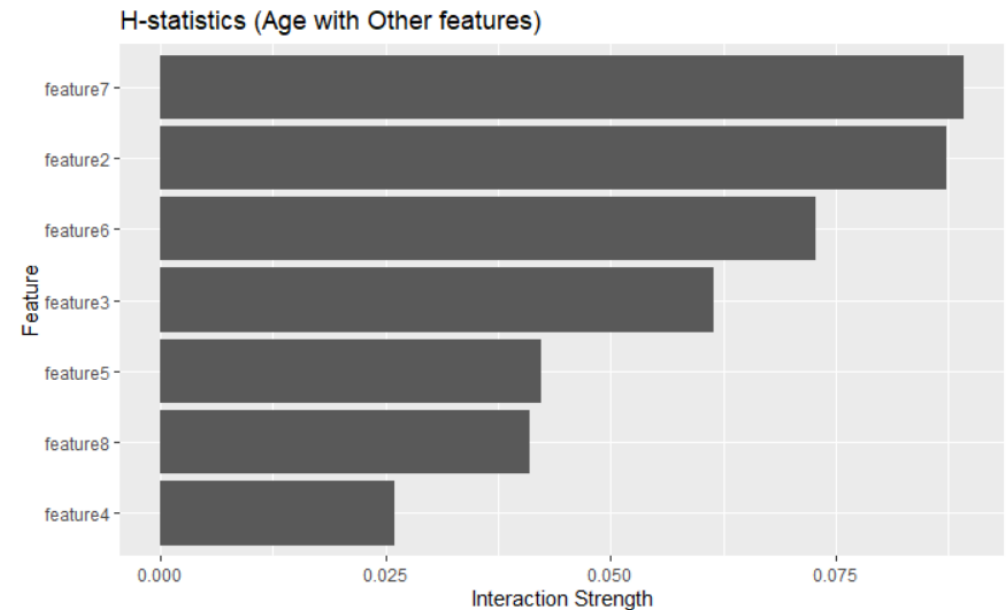
H-statistic

Using partial dependence (PD) decomposition, the interaction **H-statistic** shows the amount of variance (difference between observed PD and the no-interaction PD) explained by the interaction

- 0 means there is no interaction between 2 features.
- 1 means the prediction only depends on the interaction.

Interpretation:

Here, the H-statistics shows that *Feature7* has the largest interaction strength with *Age* so this pair is chosen to plot the PDP interaction effect.



Section 2: Sample case study

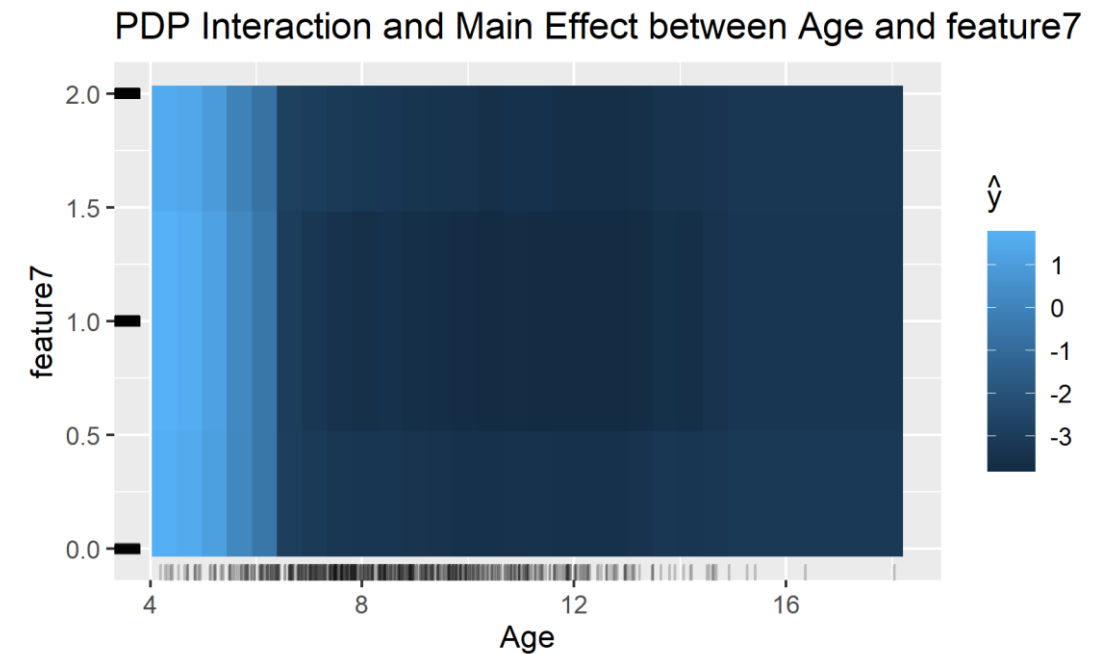
PDP plot (interaction effect)

■ Interpretation:

- PDP does not show obvious interaction pattern between *Age* and *Feature7*. The prediction is influenced mostly by Age value.
 - When Age is below ~7, the predictions are above -2 on average no matter what *Feature7* value is.
 - When Age is above ~7, the predictions are below -2 on average.

■ Potential usage:

- It assists in understanding the pattern in the interaction effect between two features.
- It helps to identify the features that are low in main effect but contribute a lot when they combine with another feature.



Section 2: Sample case study

Shapley values (global)

SHAP tries to fairly distribute payouts among the features to get the marginal contribution of a feature to the prediction.

■ **What it does:**

- Measure contribution of each feature in predicting one observation
- Rank features by their average contribution to predictions

■ **Intuition:**

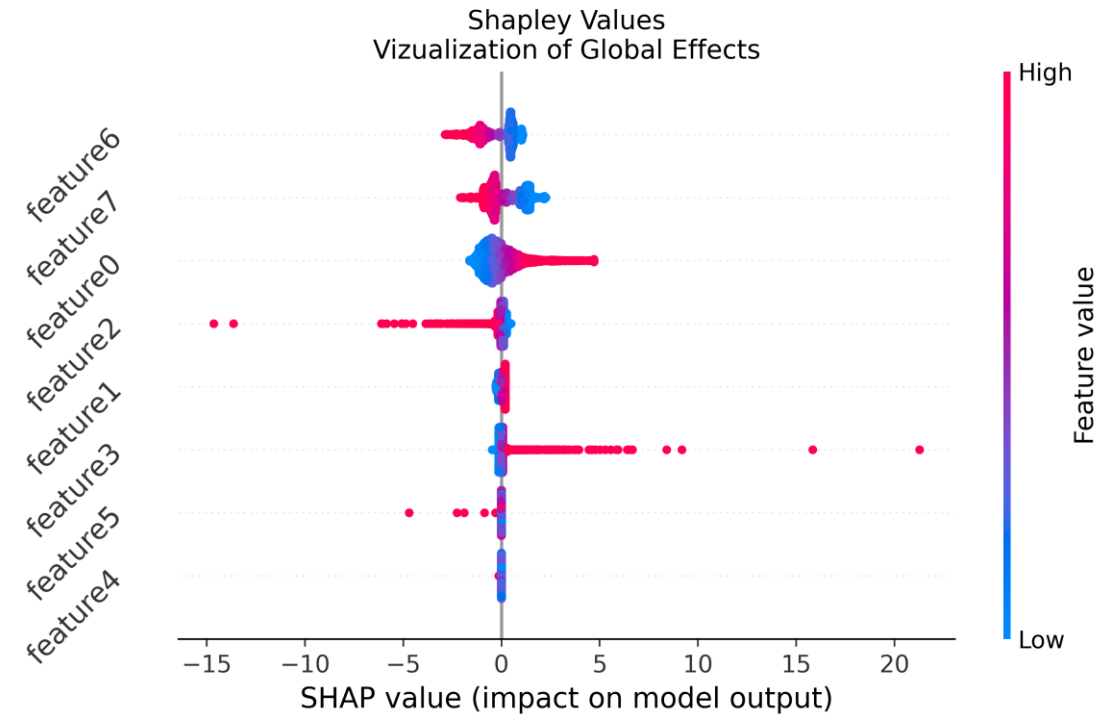
- Inspired by a method from coalitional game theory that can fairly distribute the payout among players.
- Assume that predicting the outcome of an observation is a game, where each feature is a player, and the prediction outcome is the payout.

Advantage

- Solid theoretical foundation
- Difference between prediction and average prediction is fairly distributed among features

Limitations

- Computationally expensive
- Need access to data
- Suffer from unrealistic data instance



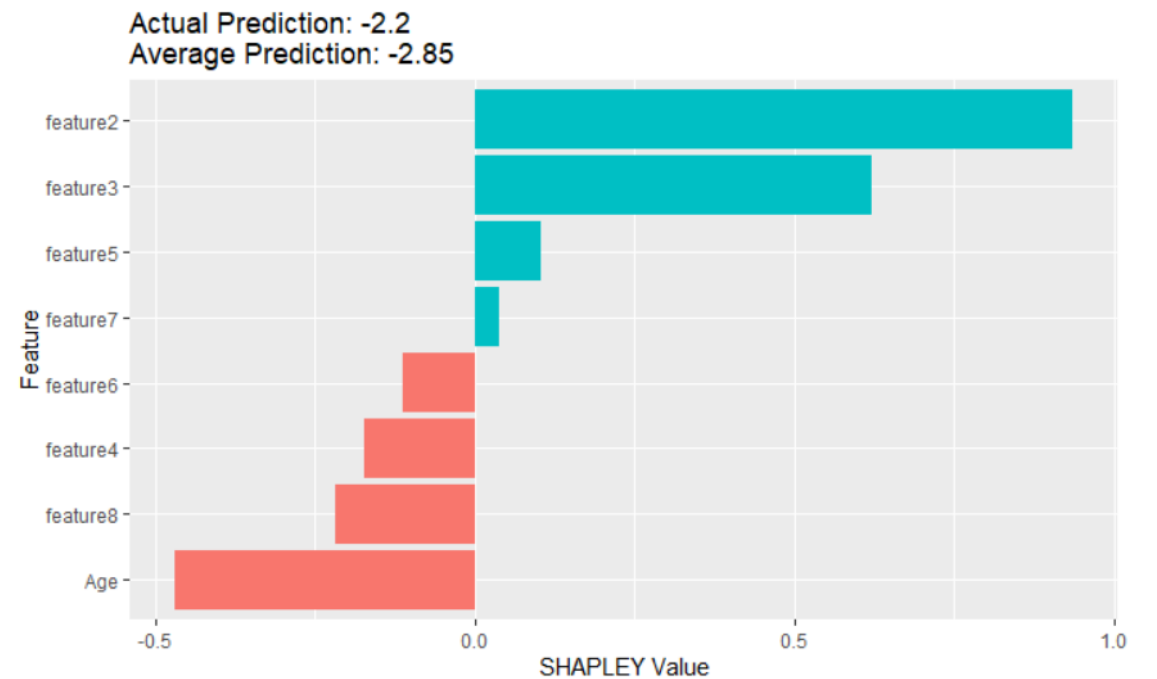
Section 2: Sample case study

Shapley values (local)

■ **Interpretation:**

- The actual prediction value is -2.2 compared to the average prediction value -2.85.
- *Feature2* marginally contributes 0.94 to the prediction on average.
- Contribution of a feature is like $\beta \cdot \text{value}$ in linear regression

Feature	Value	Contribution	Current value
[start]	-	-2.85	-2.85
feature2	0.368	0.94	$= -2.85 + 0.94 = \mathbf{-1.91}$
feature3	2.646	0.62	$= -1.91 + 0.62 = \mathbf{-1.29}$
...	
feature8	0.0	-0.22	$= -1.49 - 0.22 = \mathbf{-1.73}$
Age	12.137	-0.47	$= -1.73 - 0.47 = \mathbf{-2.2}$



Section 3: Final thoughts

- There exist many more methods for ML model interpretability
- Statistical significance of effects can also be tested with more advanced methods
- Packages are available in most languages (R, Python, etc.)

Thank you

It's Time for a Poll!

Did you find the content of this session helpful?

- Yes
- No

**Advance to next slide
for the poll**

It's Time for a Poll!

Did the content of this presentation relate to your own work?

- Yes
- No

**Advance to next slide
for the poll**

It's Time for a Poll!

Which topic would you like to learn more about?

- Solutions to advance multi-center studies in China
- National Longitudinal Cohort of Hematological Diseases in China (NICHE)
- Applications of ML-powered dynamic disease models on existing data
- Available data resources in China (e.g., HSAS)
- Interpretability of machine learning
- None of those

**Advance to next slide
for the poll**

It's Time for a Poll!

Would you be interested in collaborating with us on one of the following research topics?

- Solutions to advance multi-center studies in China
- National Longitudinal Cohort of Hematological Diseases in China (NICHE)
- Applications of ML-powered dynamic disease models on existing data
- Available data resources in China (e.g., HSAS)
- Interpretability of machine learning
- None of those

**Advance to next slide
for the poll**