

Workshop

How artificial intelligence will reshape market access

Samuel Aballéa, COO, InovIntell

Daisuke Goto, Director, Outcomes Research, Organon & Co.

Lee Moore, Executive Director, Global Value & Access, Gilead Sciences

Agenda

- Introduction
- Use cases
 - HAS transcripts mining
 - Synthetic data generation
 - Application in the Covid-19 pandemic
- Discussion

Polling Instructions

Scan QR Code on Screen with Phone Camera



Or visit **menti.com** on your phone browser and type the following code: **7631 2521**

Poll – Experience with AI



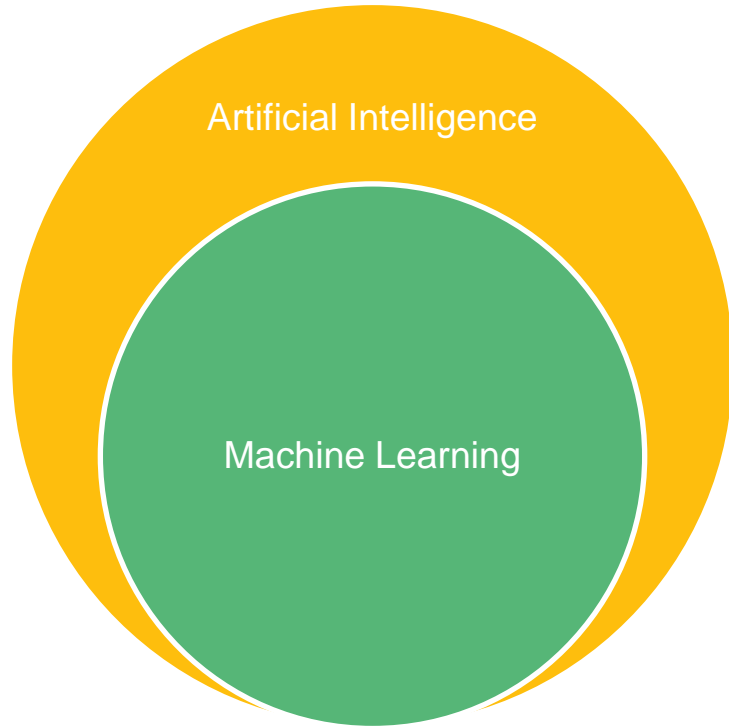
- Which statement do you most agree with?
 - I am well familiar with AI, it is a pivotal technology for some of my projects.
 - I have tried using AI technology for some projects, but my understanding of its potential is still limited.
 - I have used some standard AI applications (e.g. automatized translation, ChatGPT), but nothing more.
 - I am not using AI at all.

Poll – Sentiment towards AI



- Do you think that the potential of AI to help you meeting your objectives, as a market access professional is...?
 - Large
 - Moderate
 - Small

Definitions



- **Artificial intelligence**

“Artificial Intelligence (AI) refers to the ability of machines to perform tasks that typically require human-like intelligence, such as recognizing speech, making decisions, and learning from experience. ”

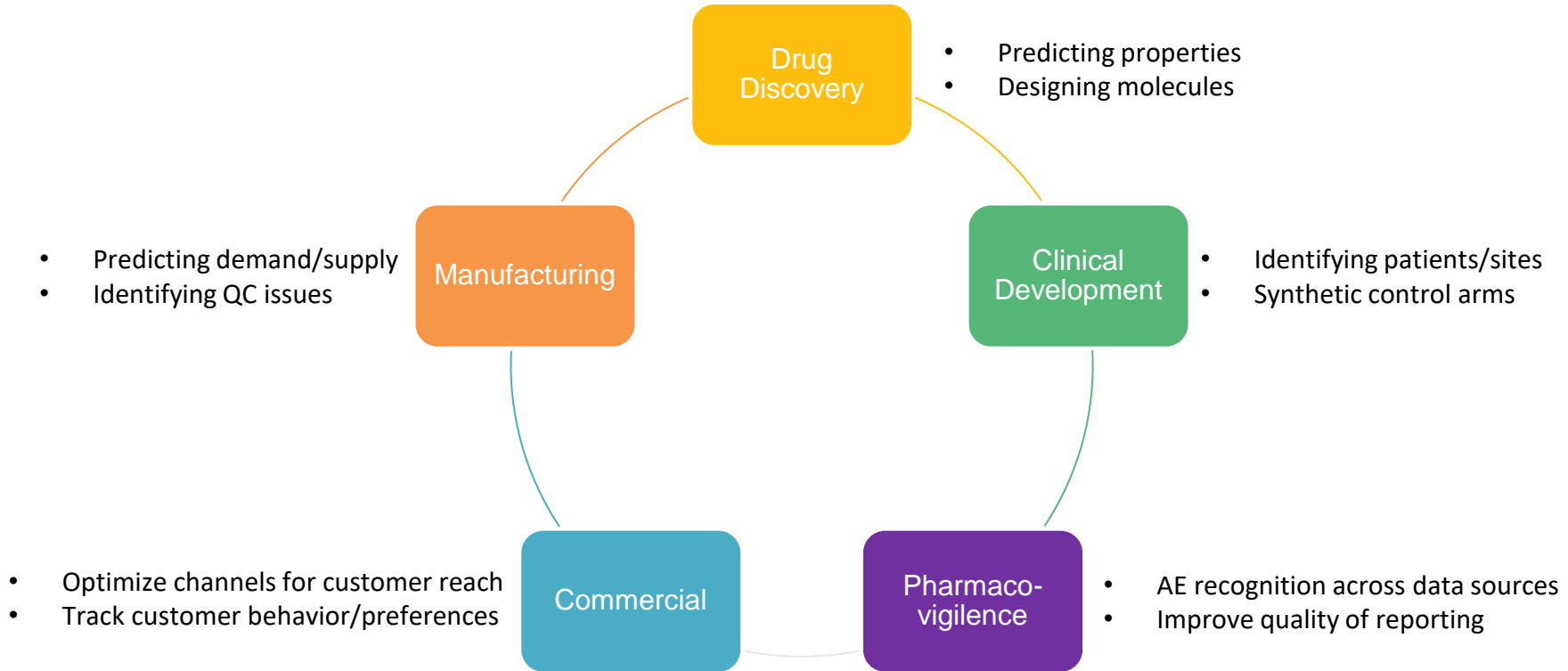
(ChatGPT, 2023)

- **Machine learning:**

“Field of study that gives computers the ability to learn without being explicitly programmed.”

(Arthur Samuel, 1959)

Current utilisations of AI in pharma



Case study:

Discovering patterns in HAS meeting transcripts with text mining

Samuel Aballéa, InovIntell



Analysis of HAS transcripts: Background

- ASMR is a key determinant of the price in France

I	II	III	IV	V
Major	Important	Moderate	Minor	Inexistent

Possible price premium

- Transcripts of the Transparency Commission meetings publicly are available.

Research question:

What are the topics associated with ASMR ratings?

Analysis of HTA transcripts: Material

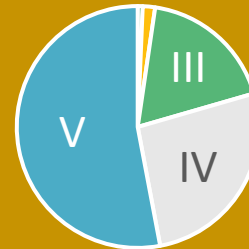
473 CT transcripts since 2018

428 CT transcripts with
unique decision

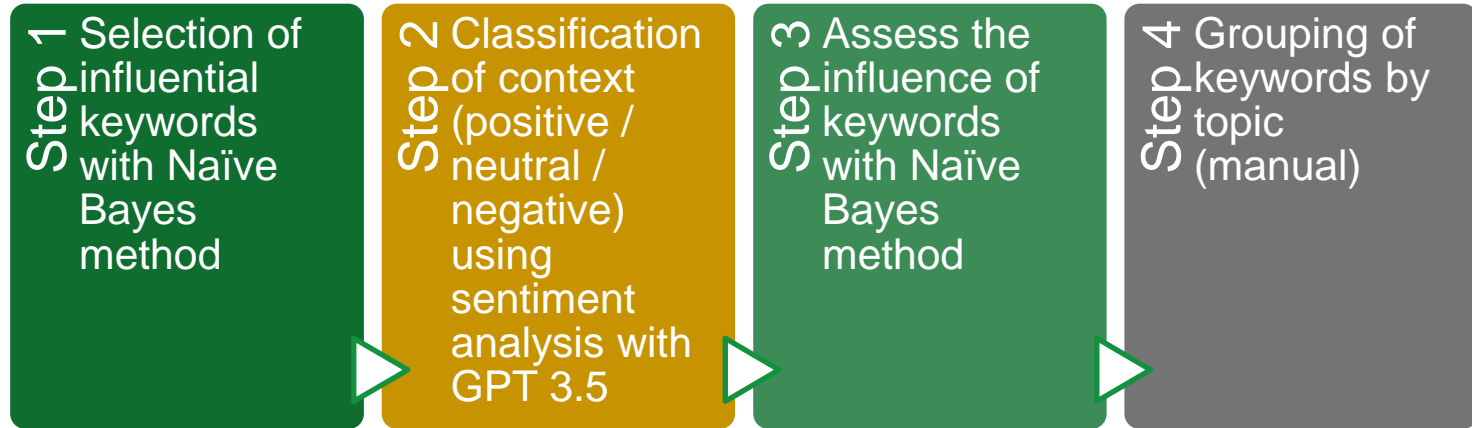
124,633 sentences

373 CT decisions

309 CT decisions



Analysis of HTA transcripts: Methods



Results: most influential keywords

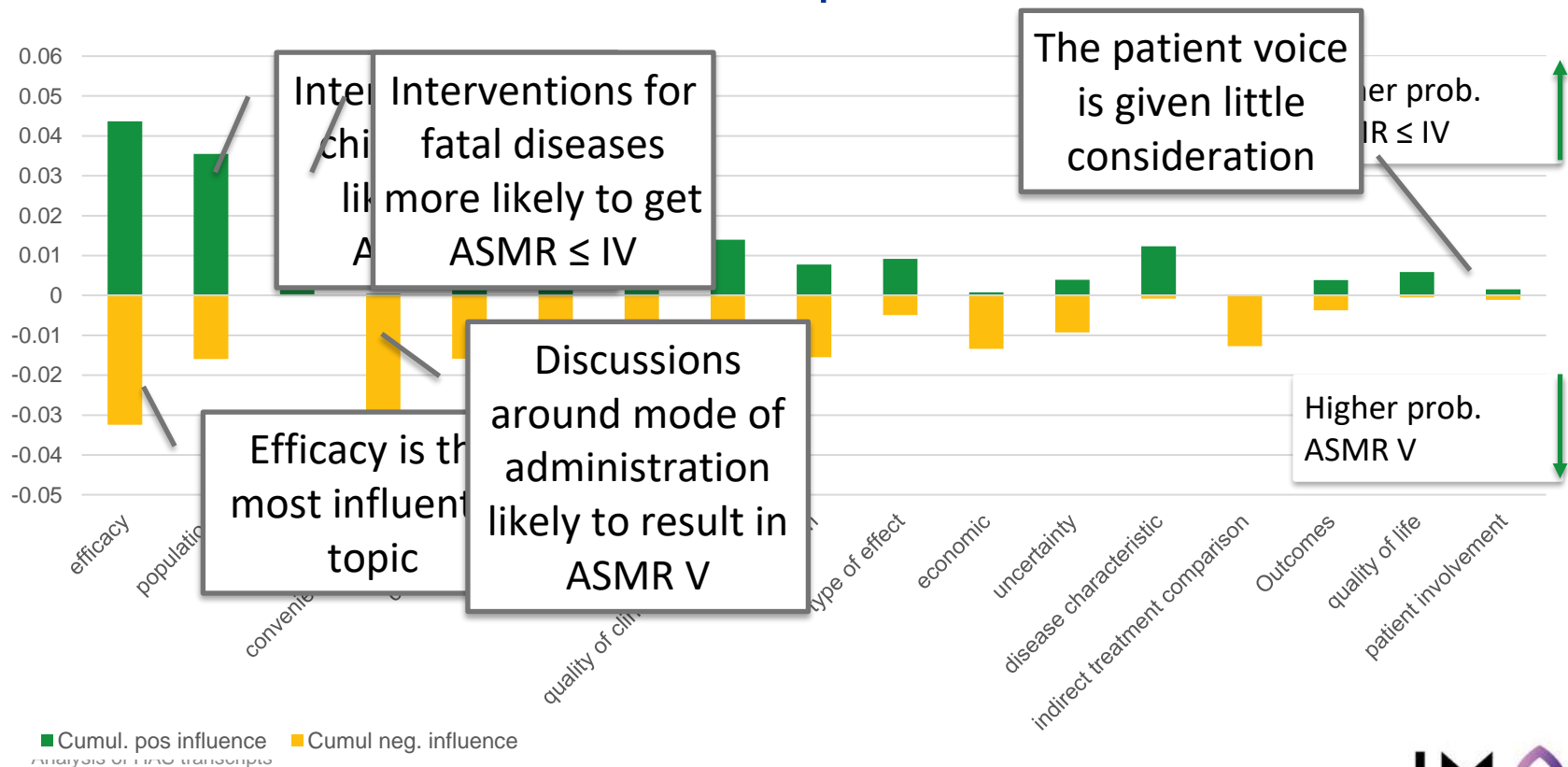
Favorable

Survival
Arm
Toxicity
Relapse
Combination
Gain
Mortality
Infant
Quality of life

Unfavorable

Injection
Meta-analysis
Placebo
Dose
Non-inferiority
Oral
Responder
Comparator

Results: influence of topics



Conclusion

Key takeaways

- We can gather insights from a large corpus of documents with **manageable effort** using AI.
- This approach is **less prone to researcher bias** compared to standard approaches.
- Enable selection of documents for in-depth review.

Next steps

- Release the hypothesis of independence between keywords
- Improve the sentiment classification model

Other applications

- CEESP transcripts
- HTA reports in different countries

Case study: Generating synthetic patient data



Samuel Aballéa, InovIntell

Generative models

“A market access workshop in Miami for professionals from the pharmaceutical industry.”



<https://deepai.org/machine-learning-model/text2img>

International Market Access Society (IMAS) 2023. As presented at ACCESS 2023 Annual Meeting & Expo in Miami Beach - April 3-5, 2023.

Generative models

“A market access workshop for professionals from the pharmaceutical industry, on the beach in Miami.”



<https://deepai.org/machine-learning-model/text2img>

International Market Access Society (IMAS) 2023. As presented at ACCESS 2023 Annual Meeting & Expo in Miami Beach - April 3-5, 2023.

Generative models: ChatGPT

“Make a table with data for 8 patients who visited a general practitioner including age, gender, reason for consultation in text, ICD-10 code and the prescription made (if any).”

Patient ID	Age	Gender	Reason for Consultation	ICD-10 Code	Prescription
1	35	Male	Sore throat	J02.9	Amoxicillin 500mg, Ibuprofen 400mg
2	28	Female	Lower back pain	M54.5	Naproxen 500mg, Physical therapy referral
3	50	Male	High blood pressure	I10	Losartan 50mg
4	62	Female	Annual check-up	Z00.00	None
5	42	Male	Headache	R51	Paracetamol 500mg, Rest and hydration recommendation
6	19	Female	Acne	L70.0	Tretinoin cream, Benzoyl peroxide wash recommendation
7	73	Male	Knee pain	M25.56	Ibuprofen 400mg, Referral to orthopedic specialist
8	55	Female	Menopausal symptoms	N95.0	Hormone replacement therapy

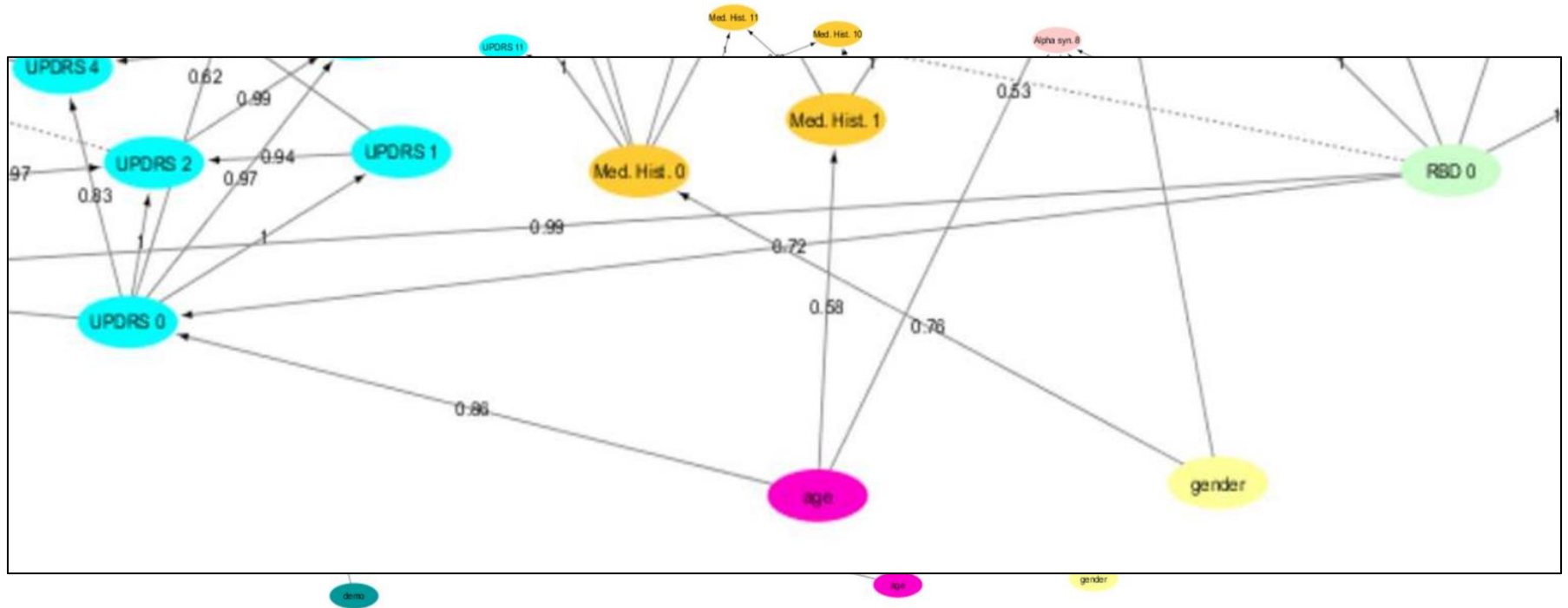
<https://chat.openai.com/chat>

Generating synthetic clinical datasets in PD

- **Variational Autoencoder Modular Bayesian Networks** (VAMBN) method, developed by Fraunhofer Institute.
- Designed for simulation of clinical study data
 - Limited sample size (100's)
 - Highly heterogeneous
 - Longitudinal
 - Presence of missing values
- Application to Parkinson's Progression Markers Initiative (PPMI) study
 - Multimodal, longitudinal observational study with data collected using standardized protocols
 - Dataset including
 - 362 de novo PD patients + 198 healthy controls
 - 266 clinical variables measured at 11 visits

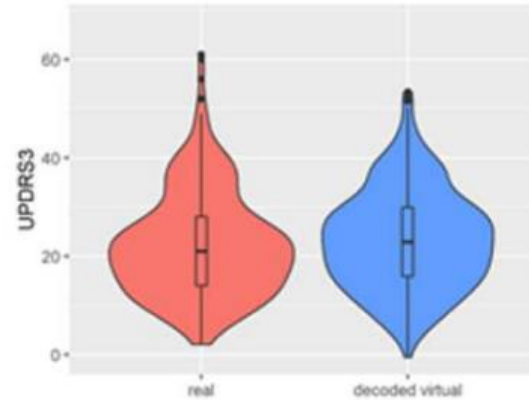
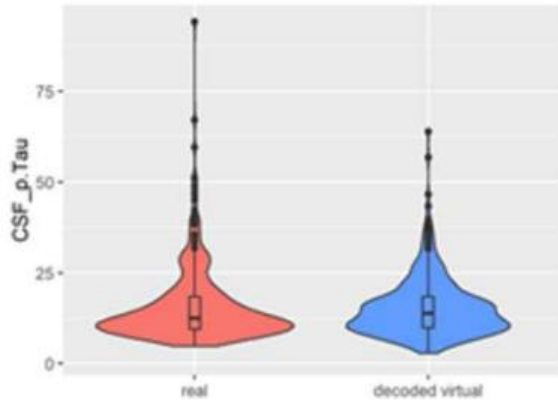
Reference: Gootjes-Dreesbach, 2020, *Front. Big Data*

Generating synthetic clinical datasets in PD



Reference: Gootjes-Dreesbach, 2020, *Front. Big Data*

Generating synthetic clinical datasets in PD



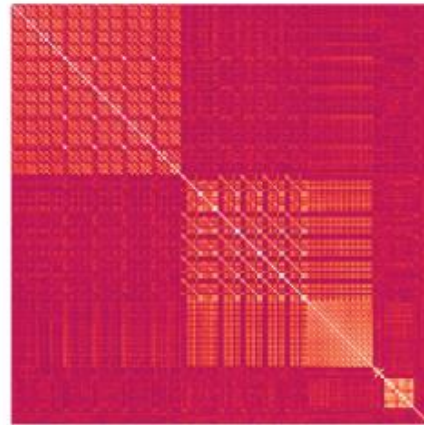
Marginal distributions of variables are similar between real and virtual data.

Source: Gootjes-Dreesbach et al., 2020, Frontiers in Big Data

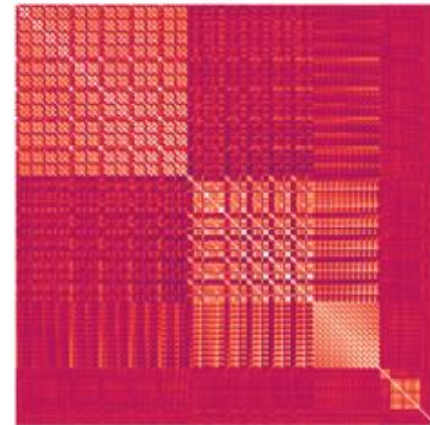
International Market Access Society (IMAS) 2023. As presented at ACCESS 2023 Annual Meeting & Expo in Miami Beach - April 3-5, 2023.

Generating synthetic clinical datasets in PD

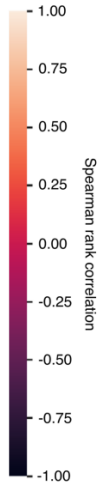
The structure of correlations between variables is similar in real and virtual datasets.



Real data



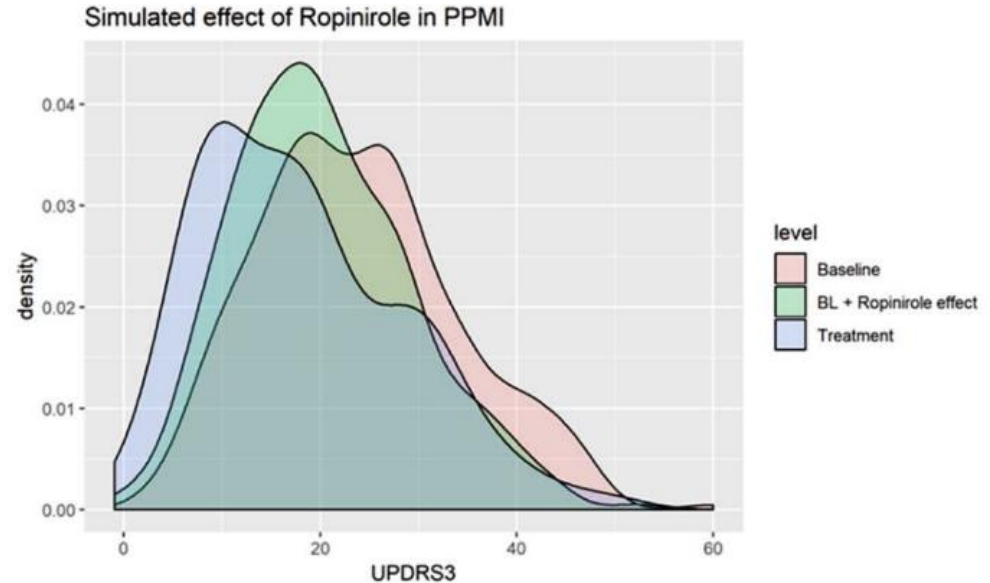
Virtual data



Source: Wendland *et al.*, 2022, npj Digital Medicine

Generating synthetic clinical datasets in PD

Simulation of ropinirole effect in PPMI patients using a combination of models based on PPMI and a clinical trial.



Source: Gootjes-Dreesbach et al., 2020, Frontiers in Big Data

Synthetic EHR dataset based on GPRD Aurum

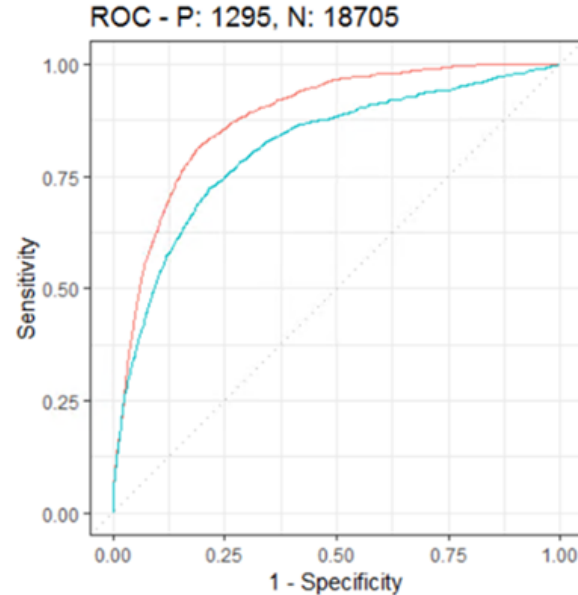
Comparison of the performance of models predicting the occurrence of stroke

Model based on
Ground Truth data:

Area under the ROC
curve= 0.812

Model based on
Synthetic data:

Area under the ROC
curve= 0.881



Reference: Tucker, 2020, npj Digital Medicine

Strengths of synthetic patient data

No data privacy concern



Simulation of counterfactual scenarios



Data augmentation



Possibility to merge virtual cohorts



Applications of synthetic data in market access



User-friendly data interrogation tools for stakeholder engagement



Synthetic control arms



Predictive modelling, extrapolation (incl. cost-effectiveness modelling)



Merging datasets with different treatment arms or different outcomes

Use cases:

Applications in the COVID-19 Pandemic



Daisuke Goto, PhD
Director
Outcomes Research
Organon & Co.

Disclaimer

This presentation was prepared or accomplished by Daisuke Goto in his personal capacity. The opinions expressed in this presentation are the author's own and do not reflect the view of his employer.

Use of AI and machine-learning algorithms for the containment of COVID-19 pandemic

The screenshot shows a Google Scholar search results page. The search query is "real-world data covid 19 ai machine-learning outcomes research". The results are sorted by relevance and show several articles. The first article is titled "A machine learning model on Real World Data for predicting progression to Acute Respiratory Distress Syndrome (ARDS) among COVID-19 patients" by N. Lazzarini et al. Other articles include "Risk Assessment of COVID-19 Cases in Emergency Departments and Clinics With the Use of Real-World Data and Artificial Intelligence: Observational Study" by E. Logarás et al., "Machine learning for COVID-19—asking the right questions" by P. Bachiger et al., "... and prognostic factors for intensive care unit admission of patients with COVID-19: retrospective study using machine learning and natural language ..." by Savana COVID-19 Research Group, and "Machine learning approaches in COVID-19 diagnosis, mortality, and severity risk prediction: A review" by N. Albaila et al.

COVID-19's contribution to AI and machine-learning

- “real-world data covid 19 ai machine-learning outcomes research” on Google Scholar returned 36,500 publications in 2019 – today (retrieved on 14 Mar 2023)
- The same search for “influenza” returned 7,950 results

Machine-learning based clinical decision support

- Lam C, Siefkas A, Zelin N, et al. Clinical Therapeutics 43(5) May 2021, pp 871 - 885
 - Data: 2364 Covid-19 patients in 10 US hospitals from EMR records
 - Objective: Identify patients in whom treatments are associated with improved survival outcomes
 - Model: Machine-learning model developed with a large number of demographic and clinical covariates
 - Findings: the machine-learning algorithm “had the capacity to identify a group of hospitalized patients with COVID-19 in whom treatment with either a corticosteroid or remdesivir was associated with a statistically significant survival benefit.”
- Gao Y, Cai GY, Fang W., et al. Nat Commun. 2020 Oct 6;11(1):5033.
 - Data: 2520 covid-19 patients (discharged or expired) from hospitals’ EMR records located in Wuhan
 - Objective: Identify patients with high mortality risks
 - Model: Machine-learning model developed with a large number of clinical covariates
 - Findings: “accurate prediction of physiological deterioration and death for COVID-19 patients up to 20 days in advance using clinical information in EHRs on admission, and validated it both internally and externally.”
- Wang G, Liu X, Shen J et al. Nat Biomed Eng. 2021 Jun 1; 5(6): 509–521.
 - Data: 145,202 chest X-ray images from China Consortium of Chest X-ray Image Investigation (CC-CXRI)
 - Objective: develop an AI system for the detection of viral pneumonia
 - Model: Application of Google’s semantic image segmentation algorithm (DeepLab v3)
 - Findings: The AI system achieved comparable performance to the senior radiologists’ level for the viral pneumonia diagnosis

Epidemiology of COVID 19

- Ayris D, Imtiaz M, Horbury K, et al. Intelligent Systems with Applications. 14:20068.
 - Data: Longitudinal COVID-19 data from Johns Hopkins
 - **Objective: Develop two novel machine-learning methods for longitudinal prediction of COVID-19 transmission**
 - Model: Two machine-learning models
 - Findings: Machine learning models are capable of accurately predicting longitudinal transmission of COVID-19 using historical data (error rates in 1.6% and 0.6% - error from the actual data)
- Naemi M, Naemi A, Ekbatani RZ, et al. Advances in Sustainability Science and Technology. 2022:107-118.
 - Data: Oxford covid-19 Government Response Tracker (including the number of cases and enforced policies; such as school closing, public events cancelation and restriction on gathering)
 - **Objective: Develop machine learning models to investigate the effect of social restrictions, including restrictions on schools, workplaces, public events, gatherings, and internal and international flights, on the control of virus spread in Denmark**
 - Model: , three machine learning models, including random forest (RF), extreme gradient boosting (XGB), and long short-term memory neural network (LSTM)
 - Findings: LSTM performed best among the three algorithms in predicting COVID-19 cases based on enforced policies

Genomics of COVID 19

- Carapito R, Li R, Helms J. et al. Science Translational Medicine. 2021;14(629)
 - Data: Genomic sequencing data from 72 critical and non-critical otherwise healthy young COVID-19 patients
 - Objective: Identify genomic factors for life-threatening worsening of COVID-19
 - Methods: Combination of multiple machine learning technics
 - Results: Identified ADAM9 as a driver of disease severity
 - **Note:** ADAM9 is one of the well-studied genes: this gene is known to be associated with progression of many types of cancer (Haoyuan MA and Yanshu LI. Cell Adhesion & Migration. 2020;14(1):165-181)

Novel approach to fight COVID-19?

- Laponogov I, Gonzalez G, Shepherd M et al. Human Genomics. 2022;15.1
 - Aim: Use of network machine learning methods to identify bioactive anti-covid-19 molecules in foods to construct a “food map” to fight COVID-19.
 - Data: Gene-protein connections data (compound and human genes) and protein- protein connections data
 - Methods: Machine-learning based human interactome networks
 - “A database of 7694 bioactive food-based molecules was run through the calibrated machine learning algorithm, which identified 52 biologically active molecules”
 - Contributors from Imperial College London, Yale and Vodafone
 - Data and algorithm are both in the public domain ([iAnalytica / corona-ci-correlator — Bitbucket](https://github.com/ImperialCollegeLondon/iAnalytica_corona-ci-correlator))
 - Do foods “cure” or “prevent” COVID-19 in the real-world settings?

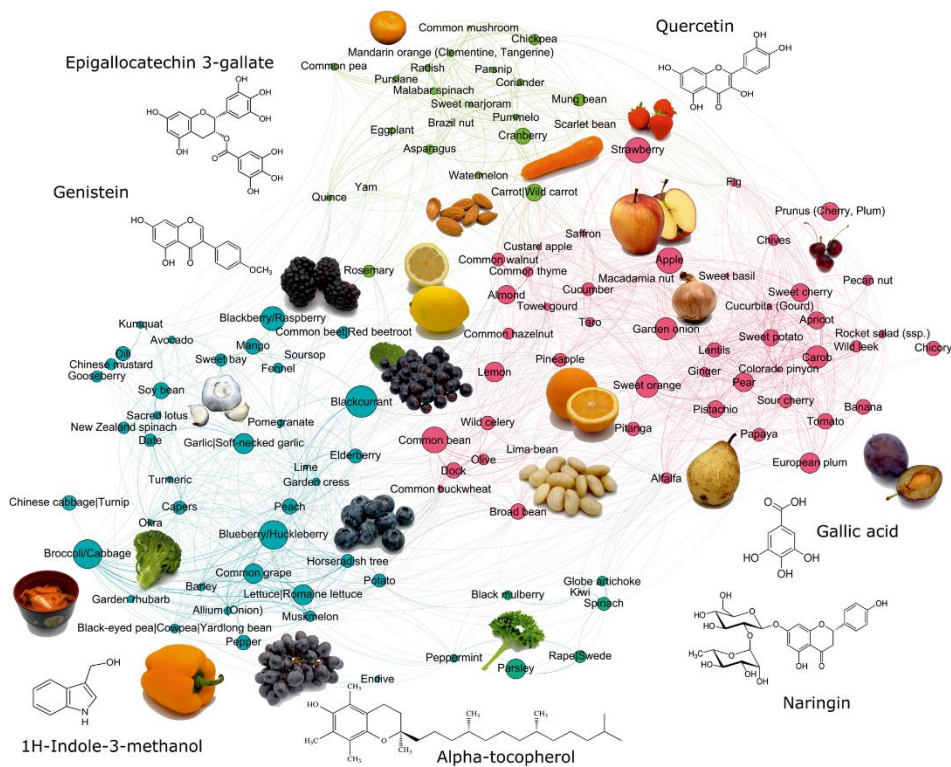


Fig 3. Laponogov I, Gonzalez G, Shepherd M et al. Human Genomics. 2022;15.1
 Reproduced under a Creative Commons Attribution 4.0 International License

Impact to Market Access

- Flood of value and clinical evidence from AI and Machine-learning assisted analysis of real-world data
 - “Hundreds of AI tools have been built to catch covid. None of them helped.” - Senior editor for AI, MIT Technology Review (MIT Tech Rev. July 30, 2021)
- Machines (without human intervention) do not understand causality, only correlations
 - “Machine-learning approaches are powered by identification of strong, but theory-free, associations in the data. Confounding makes it a substantial leap in causal inference to identify modifiable factors that will actually alter outcomes.” – Jonathan Chen and Steven Asch, Veterans Affairs Palo Alto Healthcare System (N Engl J Med. 2017;376(26): 2507-2509)
- HEOR studies need significant human intervention and AI and machine-learning often lead to additional efforts due to the lack of resources to adopt
 - Just like any conventional real-world studies, we need the following to conduct high-quality studies
 - Fit-for-purpose data
 - Fit-for-purpose statistical methods
 - Expert research scientists
- Flood of new studies makes it difficult to evaluate the quality of evidence for market access professionals and customers (healthcare systems and payors)
 - Propagation of AI and machine-learning further enhances importance of collaboration between HEOR scientists and market access leaders



**REIMAGINE
ACCESS.
DRIVE
INNOVATION.**

Summary & Discussion

Challenges for using AI in market access

- Requires human and technical resources
- Data-related
 - Unstructured, heterogenous, incomplete
 - Highly sensitive
 - Hardly accessible
- Lack of transparency
- Difficulty to communicate
- Limited understanding of causality

Current applications of AI for market access

- Gather insights from large document corpuses
- Help with literature reviews
- Help with writing, e.g. HTA dossiers, GVDs
- Data science / RWE generation
 - Identifying hurdles to patient access
 - Identifying patients eligible for treatment
 - Generating insights at patient level
 - Extracting information from unstructured medical records
- Inform pricing and reimbursement strategies, by fitting models on past decisions

Benefits of AI

- Save time / reduce costs
 - Possibility to explore large numbers of scenarios
- Systematic approaches, avoid or reduce human bias
- Improved predictions, in particular at patient level
- Capacity to handle large volumes of data
- **Opens up new possibilities**

Expected future applications

Methodological
innovation

Integration of different
tools into platforms

Data consolidation

Validation studies



- Prediction of HTA outcomes and prices
- Prediction of future landscape
- Automatisation of literature reviews
- Synthetic control arms / augmentation of clinical trials
- Causal inference
- Interactive tools or platforms for stakeholder engagement, e.g. data exploration, prediction
- Live / real-time assessment of treatments

Group discussions

- Where do you see the greatest potential for applications of AI in market access?
- What are the main hurdles against applications of AI in your company?



**REIMAGINE
ACCESS.
DRIVE
INNOVATION.**

Thank you!



**REIMAGINE
ACCESS.
DRIVE
INNOVATION.**

Back-up slides

AI/Machine-learning in HEOR vs Computer Science: differences in input data

- Computer science (face recognition, hand-writing recognition, and ChatGPT)
 - Highly structured data (such as rectangular images or videos)
 - One type of input at a time
 - Option to ignore low quality input (such as low-resolution data)
- HEOR (evidence generation)
 - Lot of surprises (missing data, different documentation standards, scanned images)
 - Diverse clinical practices (specialty-specific, payor-specific, country-specific practices)
 - Multiple inputs interacting and codependent on each other
 - Limited option to ignore low quality input (vs. bias)
 - No standardized format for many input parameter values (vs. picture formats [such as JPEG] are universal)
- Machine-learning algorithms are not mature enough to automatically generate customized new algorithms to incorporate the above-listed heterogeneity that HEOR studies need to deal with.

Machine-learning generated HEOR evidence for market access

- Analytical methods that are mature enough for application today
 - Natural language processing based
 - Literature review
 - Unstructured data (EMR) analysis
 - Large structured data analysis (“big data”)
 - Predictive analysis
 - Patient clustering
 - Personalized medicine
- Future analytical applications that do not have concrete methodological standards (*)
 - Causal inference
 - Economic modeling for health technology assessment

* Based on ISPOR REPORT. Value in Health. 25(7), pp. 1063-1080, 2022

Current uses: Literature reviews

- Pre-requisite for success: fit-for-purpose NLP algorithm and validation
- Value to payors and providers: availability of latest evidence to aid their decisions
- Current uses
 - Help researchers to filter out irrelevant studies
 - Help researchers to curate data from the literature
- Challenges:
 - Newly found studies (and machine-curated data) typically require human review
 - Experts still need interpret study results and generate clinical insights (such as impact to clinical guidelines)
- Future use:
 - Automated systematic literature reviews
 - Automated system to suggest new guidelines and policies based on new data

Current uses: Analysis of unstructured (text) data from EMR

- Pre-requisite for success: comprehensive understanding of the clinical practice and limitations and potential biases to develop a customized fit-for-purpose algorithm
- Value to payors and providers: real-time insights from real-world clinical practices
 - Insights include safety, effectiveness, adherence to standard of care (quality control)
- Challenges:
 - Low quality data
 - Missing data not at random
 - Patient privacy - some data (such as images and genomic data) are highly sensitive
- Future:
 - Use of real-time natural language processing to aid clinicians to enter all necessary data for research purposes
 - Automated evidence generation

Current uses: Predictive HEOR analysis (incl. obtaining inputs for models)

- Pre-requisite for success: large, longitudinal datasets
- Value to payors and providers: prediction of clinical and economic outcomes when there are a large number of factors in play
 - Prediction of outcomes in performance-based (outcomes-based) reimbursement
 - No need to identify complex causal relationships
- Challenges:
 - Correlations do not mean causality
 - Algorithms might not work well with underrepresented groups
 - Highly dependent on the quality of the data
 - Data access might be limited and require patient consent
 - Highly sensitive data from clinical sources (such as genomic data) and non-clinical sources might be needed
- Future:
 - Integration into decision support systems
 - Wider-use of performance-based reimbursement

Current uses: Patient clustering

- Pre-requisite for success: large data
- Value to payors and providers: identify a group of patients which requires special attention from payors and providers
 - Based on a well-defined objective of clustering, machines learning algorithm creates subgroups
- Challenges:
 - Highly dependent on the quality of the data
 - Highly sensitive data from non-clinical sources might be needed
 - Data access might be limited and require patient consent
- Future:
 - Automatic identification of sub-populations of patients whose outcomes differ from the other populations – aiding clinical and reimbursement decision making

Current uses: Personalized medicine

- Pre-requisite for success: large reliable genomic sequencing data
- Value to payors and providers: identify effectiveness of personalized rare disease treatment for each patient
 - Ability to select the most effective treatment for each patient
 - Possibility to tie expected outcomes with reimbursement
- Challenges:
 - Personalized medicine is evolving daily – we cannot solely rely on data available today.
 - Scientific mechanisms are often poorly understood (compared to many non-rare disease treatments)
 - Individual data are highly sensitive and subject to discussion in bioethics
- Future:
 - Automatic identification of best personalized treatment

Future uses: Causal inference

- Why not mature enough today?
 - Statistical methods (on non-interventional data) are not mature enough to replace interventional studies, although very careful use of statistics and high-quality dataset might be considered “good enough” by some regulators (many regulatory and HTA agencies are learning and developing guidelines)
 - Machine learning (AI) technologies might be considered in the future
- Value to payors and providers: evidence with high internal validity (scientific validity of efficacy in the targeted population) from real-world data (faster and cheaper)
- Challenges:
 - Must gain consensus on the scientific methods and standards
 - Highly dependent on the quality of the data; thus improvement of data quality will be required
 - Highly sensitive data
 - Data access might be limited and require patient consent
- Future:
 - Routine use of highly generalizable real-world evidence from non-interventional studies in regulatory and reimbursement decision making

Future uses: Economic modelling for HTA

- Why not mature enough today
 - Disease-specific economic models require disease and healthcare-system specific insights and input parameters; payer requirements must be considered
 - AI or machine-learning approaches tend to be model-agnostic and this practice is not common in today's HECON research.
- Value to payors and providers:
 - Automated economic evaluation without waiting for model development efforts by experts
- Challenges:
 - Economic assessment without models developed by health economists is currently not a norm
 - Health economic models using Excel provide some transparency and predictability to the audience
 - Highly dependent on the quality of the data
- Future:
 - Acceptance of machine-learning generated health economic analysis and continuous update of economic assessments

Ending Remarks

- AI and machine-learning are currently utilized in many HEOR efforts
 - The large majority of HEOR studies still relies on the traditional methods
 - Traditional methods are effective
 - HEOR studies need significant human intervention and AI and machine-learning often lead to additional efforts due to the lack of resources to adopt
 - In computer sciences, many tested and accepted machine-learning packages are available
 - AI and machine-learning approaches from other industries (computer science) will continue to change the way real-world evidence is generated in the pharmaceutical industry
 - Changes can be very fast



THANK YOU!