Artificial intelligence (AI) and machine learning (ML) approaches to identify and appraise registries and relevant data

9th July 2024

Housekeeping

All participants will automatically be muted.

The chat box is disabled.

Please add your questions in the Q&A box. They will be answered either during the webinar or at the Q&A session at the end.

This webinar is being recorded; we will share the recording with all registrants after the webinar.

Peter Mol

Introduction

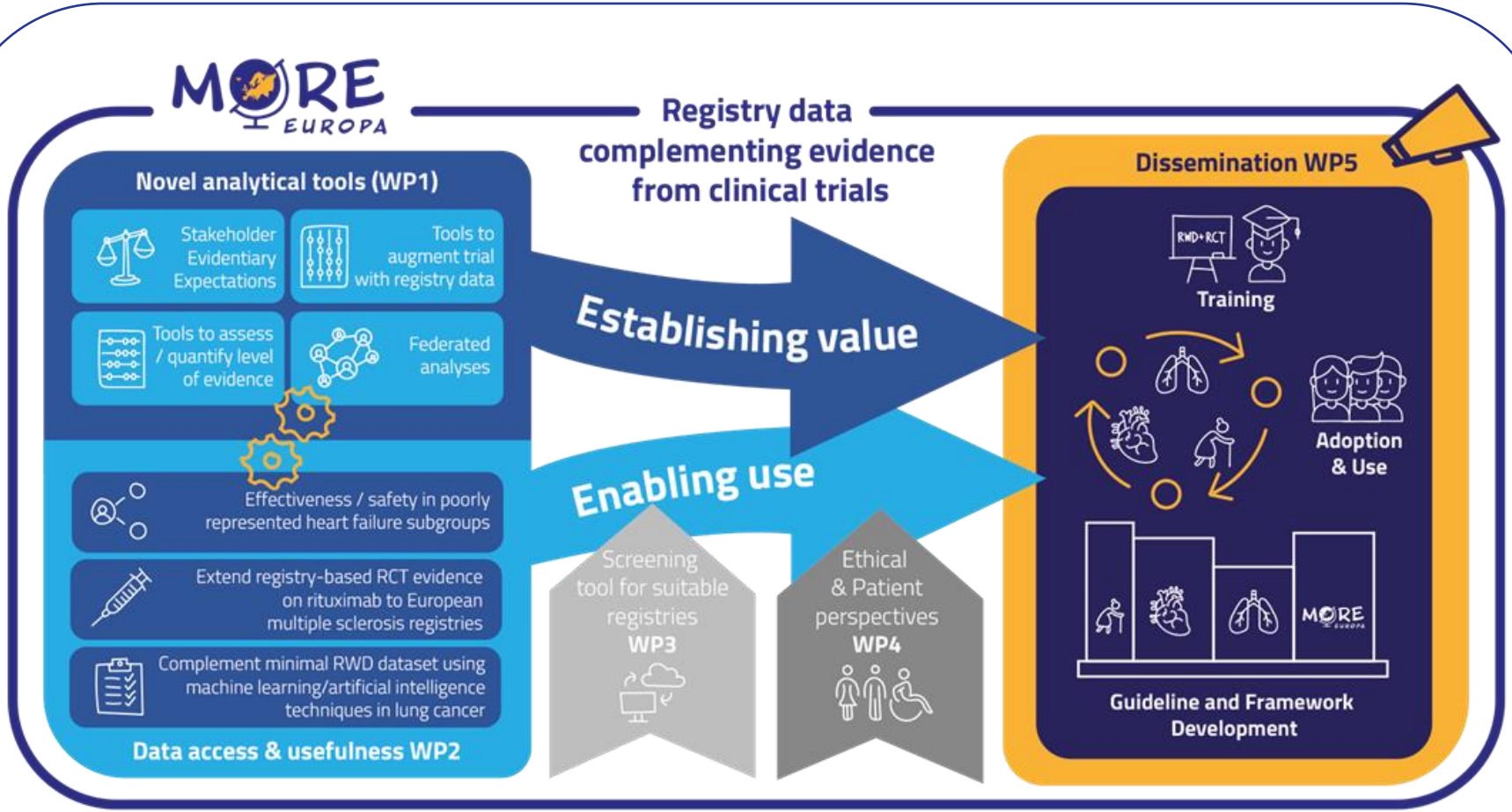
More-EUROPA

This project has received funding from the European Union's Horizon Europe Research and Innovation Actions under grant no. 101095479 (More-EUROPA). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union nor the granting authority. Neither the European Union nor the granting authority can be held responsible for them.

- Horizon-HLTH-2022-TOOL-11-02

Aims

- Establish value of registry-based RWD in augmenting RCTs
- Enable **more** **e**ffective and ethical **u**se of **r**egistry data to supp**o**rt **pa**tient-centered regulatory and health technology assessment decision-making


MORE EUROPA
Registry data complementing evidence from clinical trials
Novel analytical tools (WP1)
Stakeholder Evidentiary Expectations
Tools to augment trial with registry data
Tools to assess / quantify level of evidence
Federated analyses
Effectiveness / safety in poorly represented heart failure subgroups
Extend registry-based RCT evidence on rituximab to European multiple sclerosis registries
Complement minimal RWD dataset using machine learning/artificial intelligence techniques in lung cancer
Data access & usefulness WP2
Establishing value
Enabling use
Screening tool for suitable registries
WP3
Ethical & Patient perspectives
WP4
Dissemination WP5
RWD+RCT
Training
Adoption & Use
MORE
Guideline and Framework Development

More-EUROPA webinars

- Use of registries in regulatory decision-making
- How to appraise and apply AI and/or ML approaches to identify relevant registries, additional data and determine relevant confounders and outcomes
- How RWD elements can augment clinical trial performance, and how to use registries as a source for external comparator data
- Methods to access registry data, data sharing (consent and privacy), and federated analyses
- How to evaluate and design registry-based RCTs

More-EUROPA webinars

- ~~Use of registries in regulatory decision-making~~ (see website)
- How to appraise and apply AI and/or ML approaches to identify relevant registries, additional data and determine relevant confounders and outcomes

Next topics:

- How RWD elements can augment clinical trial performance, and how to use registries as a source for external comparator data
- Methods to access registry data, data sharing (consent and privacy), and federated analyses
- How to evaluate and design registry-based RCTs

Today

- *Learn how AI and ML can help organisations in optimising their processes*
- *Know how AI and ML can help regulators and HTAs in their decision-making*
- *Learn about the AI-driven tool developed in the context of More-EUROPA*

Topic	Speaker(s)	Duration
Enabling the safe and responsible use of AI in medicines lifecycle and regulatory science	Luis Pinheiro (EMA)	20 min
Regulator-in-the-loop: how to integrate new ML based technologies in the regulatory landscape	Nicolás Pérez González (Swissmedic)	20 min
An AI-powered tool for identification, and pre-assessment, and selection of registries to support regulatory and HTA decision processes	Billy Amzal, Ghinwa Hayek, Pascal Godbillot, Coriande Clemente (Quinten Health)	20 min
Q&A session	Moderator: Cécile Olivier	45 min
Wrap-up and conclusion	Peter Mol	5 min

Webinar Speakers

Peter Mol

Luis Pinheiro

Nicolas Perez Gonzalez

Billy Amzal, Ghinwa Hayek, Pascal Godbillot, Coriande Clemente

Cécile Ollivier

Speakers Bio

Luis Pinheiro

Luis Pinheiro is a Senior Epidemiology Expert at the European Medicines Agency (EMA), in the Data Analytics Workstream of the Data Analytics and Methods Taskforce. He is responsible for designing and conducting real-world data analytics studies and is tasked with methods development and learning within the team. Luis is also a co-chair of the AI Technical Group at EMA. Before moving to the Data Analytics Workstream, Luis worked eight years in drug safety at EMA.

Prior to joining the EMA, he was the Head of the Signal Detection Unit at Infarmed, the Portuguese Regulatory Authority and the Portuguese representative at the Pharmacovigilance Working Party at EMA. Before that, he worked in the Pharmacovigilance Unit of the Faculty of Pharmacy of Lisbon for eight years where he assisted in the teaching of Pharmacoepidemiology and Public health in undergraduate and postgraduate courses.

Luis holds a PharmD from the Faculty of Pharmacy of Lisbon (2003) and received his Masters in Epidemiology from the Faculty of Medicine, New University of Lisbon in 2008. He is pursuing a PhD in Health Data Science from ErasmusMC.

Speakers Bio

Nicolas **Löffler-Perez, PhD**

Data Scientist at Swissmedic. After obtaining his Ph.D at The Johns Hopkins University, he moved to Switzerland to work on Clinical Machine Learning in the UniversitätsSpital Zürich and the University of Zürich. While there he developed a multimodal model for comprehensive analysis of images and reports in the hospital, as well as optimisation algorithms and a number of other exciting projects.

Within Swissmedic Nicolas' primary focus is in innovation and implementation of new technologies. He currently works with the Swissmedic 4.0 team as the technical lead in multiple AI-based applications, including the search of illegal medicinal substances in e-commerce, medicament recognition from images, and document similarity.

Speakers Bio

Billy Amzal, MSc, PhD

Billy Amzal graduated from Ecole Polytechnique, from AgroParisTech and holds a PhD in Decision Mathematics from Paris-Dauphine University which was awarded by the International Society of Bayesian Analysis and by the International Biometrics Society. Over the past 20 years, Billy has developed quantitative methodologies to inform and support strategic decision making in healthcare. Prior to joining Certara, Billy led the model-based drug development initiative at Novartis Pharma. Subsequently, he developed assessment methodologies at the European Food Safety Authority. Billy was also Director of the Data Center at the NIH-sponsored Program for HIV Prevention and Treatment in Thailand. Billy has then been Senior VP at Certara in charge of Real world data and decision analytics. Billy is now leading the Quinten Healthcare business.
He acts as statistical or modelling expert for public Health Authorities (EFSA, ANSES, WHO, Gates Foundation).
Billy led hundreds of RW data projects, authored more than 100 publications, defended or submitted hundreds of regulatory and reimbursements dossiers, and led hundreds of data modelling or machine learning projects.

Pascal Godbillot, MSc, D. Eng

Pascal Godbillot is a Senior Data Scientist at Quinten Health. For the last six years he has been working on the development and deployment of machine learning solutions with a recent focus on real-world evidence. Prior to joining Quinten Health, he worked in cybersecurity on digital threat detection using various natural language processing (NLP) applications. Pascal graduated from the Ecole Centrale de Lyon and holds an engineer degree and a master's degree in data science.

Coriande Clemente, MSc, D. Eng

With an engineer degree in data and applied mathematics from "l'école des Mines de Nancy" and a MSc degree in Cognitive Science (University of Lorraine), Coriande masters statistics and machine learning tools. At Quinten Health, Coriande has developed extensive experience with real-world data through topics such as cost-effectiveness analysis, rare disease diagnosis algorithms development and synthetic data generation. She also manages NLP through problems such as extracting information from patient testimonials or systematic literature reviews.

Ghinwa Hayek, MPH

Ghinwa has an Master of Public Health in epidemiology and biostatistics from the American University of Beirut, and a Masters in Research Methods and Biostatistics for Biomedical Research from the University of Paris Saclay. After 7 years working in the public health, healthcare, and humanitarian fields in the roles of epidemiologist, and monitoring and evaluation coordinator, covering different topics covered such as mental health, maternal and newborn health, diabetes, tuberculosis, digital therapeutics, and COVID-19, she joined Quinten Health in her current role of Data Science Project Manager to bring in her technical and managerial expertise to the team.

Speakers Bio

Cécile Olivier

Data Cécile is an health engineer by training, with a Master Degree in International Regulatory Affairs. She is VP of Global Affairs at C-Path and prior to that, she was in a medtech company developing digital endpoints and for 11 years at the European Medicines Agency.

EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

Enabling the safe and responsible use of AI in medicines lifecycle and regulatory science

A regulatory perspective on AI

More-Europa

Luis Pinheiro, Epidemiologist and Data Scientist, European Medicines Agency

The Promise of AI | AI promise across the medicine lifecycle

Greater efficiency /productivity

- Automate processes
- Address scalability issues
- Leverage personal assistants (chatbots)

Do more things, faster

Reduce error

- Facilitate access to information
- Reduce human cognitive load

Provide information at the fingertip

Expose data

- Transform text data into structured data
- Reduce dimensionality of data
- Imputation of missing data

Structure and summarise information

Expand insights

- Probabilistic phenotyping
- Synthetic control arms
- Clinical prediction modelling
- Confounding adjustment
- Digital endpoints
- Heterogeneity of treatment effects

Predict probability of events

AI in Healthcare | Regulatory Drivers


Regulatory submissions


Process Analytics


Healthcare Analytics

Enabling the safe and responsible use of AI in medicines lifecycle and regulatory science

AI in Healthcare | Regulatory Drivers

Regulatory submissions

Regulate applications of AI in medicines to help create value for public health

Process Analytics

Healthcare Analytics

Enabling the safe and responsible use of AI in medicines lifecycle and regulatory science

AI in Healthcare | Regulatory Drivers

Regulatory submissions

Process Analytics

Improve **efficiency** by automating and digitalizing processes

Healthcare Analytics

Enabling the safe and responsible use of AI in medicines lifecycle and regulatory science

AI in Healthcare | Regulatory Drivers

Regulatory submissions

Process Analytics

Healthcare Analytics

Structure information and increase insights into data to inform decision-making

Enabling the safe and responsible use of AI in medicines lifecycle and regulatory science

Internal Drivers

Process analytics and healthcare data analytics

AI at EMA | Set up

Digital Innovation Lab (DigiLab)

Established 2021 to accelerate digital innovation and experimentation

Analytics Centre of Excellence

Explore analytics, AI, ML and robotics to **build pragmatic solutions and experiment with new technology**

-- Process analytics --

Health Data Lab (pilot)

Develop innovative analytics to **extract insights from healthcare data**, to provide answers/recommendations and maximise intelligence

-- Healthcare data analytics --

Infrastructure

Information Processing and Analytics

Improve the organisation's performance through streamlining information processing across the regulatory lifecycle.

-- Data analytics tools and platforms --

Enabling the safe and responsible use of AI in medicines lifecycle and regulatory science

AI experimentation | Examples

Literature screening

Use of LLMs to identify abstracts with information on the occurrence of adverse reactions

Extraction of ADRs from SmPCs

(semi) Automated extraction of ADRs from product information

Automated adjudication of ADRs

Automated adjudication of ADRs

Improve signal detection with previous reviews

Using signal reviews to improve signal detection

AI experimentation | Examples

Literature screening

Use of LLMs to identify abstracts with information on the occurrence of adverse reactions

Extraction of ADRs from SmPCs

(semi) Automated extraction of ADRs from product information

Automated adjudication of ADRs

Automated adjudication of ADRs

Improve signal detection with previous reviews

Using signal reviews to improve signal detection

AI experimentation | Examples

Literature screening

Use of LLMs to identify abstracts with information on the occurrence of adverse reactions

Extraction of ADRs from SmPCs

(semi) Automated extraction of ADRs from product information

Automated adjudication of ADRs

Automated adjudication of ADRs

Improve signal detection with previous reviews

Using signal reviews to improve signal detection

AI experimentation | Examples

Literature screening

Use of LLMs to identify abstracts with information on the occurrence of adverse reactions

Extraction of ADRs from SmPCs

(semi) Automated extraction of ADRs from product information

Automated adjudication of ADRs

Automated adjudication of ADRs

Improve signal detection with previous reviews

Using signal reviews to improve signal detection

External Drivers

Regulate AI

Guidance | AI Reflection Paper

EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

13 July 2023
EMA/CHMP/CVMP/83833/2023
Committee for Medicinal Products for Human Use (CHMP)
Committee for Medicinal Products for Veterinary Use (CVMP)

Reflection paper on the use of Artificial Intelligence (AI) in the medicinal product lifecycle

Draft

Draft agreed by Committee for Medicinal Products for Human Use (CHMP) Methodology Working Party	July 2023
Draft adopted by CVMP for release for consultation	13 July 2023
Draft adopted by CHMP for release for consultation	10 July 2023
Start of public consultation	19 July 2023
End of consultation (*deadline for comments*)	31 December 2023

Comments should be provided using this EUSurvey form. For any technical issues, please contact the EUSurvey Support.

Keywords	*Artificial intelligence, AI, machine learning, ML, regulatory, medicine, human medicinal product, veterinary medicinal product*

Reflection paper on the use of AI in the medicinal product lifecycle

Table of contents

Enabling the safe and responsible use of AI in medicines lifecycle and regulatory science

Guidance | AI Reflection Paper

General considerations & risk

Reflection paper on the use of AI in the medicinal product lifecycle

Table of contents

- AI and ML tools can effectively support the acquisition, transformation, analysis, and interpretation of data within the medicinal product lifecycle.
- A risk-based approach for development, deployment and monitoring of AI and ML tools allows developers to pro-actively define risks to be managed throughout the AI lifecycle.
- The degree of risk may depend on
  - the AI technology,
  - the context of use,
  - the degree of influence the AI technology exerts and
  - may vary throughout the lifecycle of the AI-system.
- MAA/MAHs planning to deploy AI/ML technology are expected to consider and systematically manage relevant risks from early development to decommissioning.

Bringing it all together

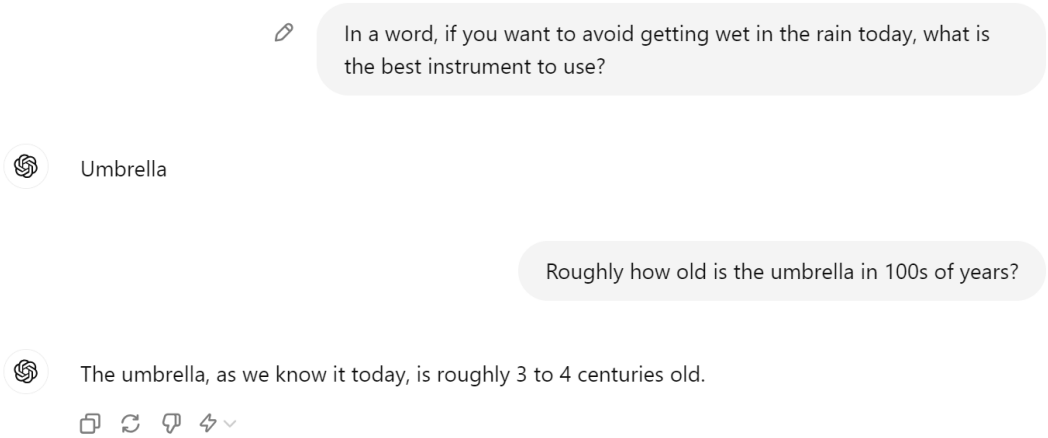
Enabling responsible and safe use of AI for the benefit of public and animal health

Safe and responsible use | AI can be wrong

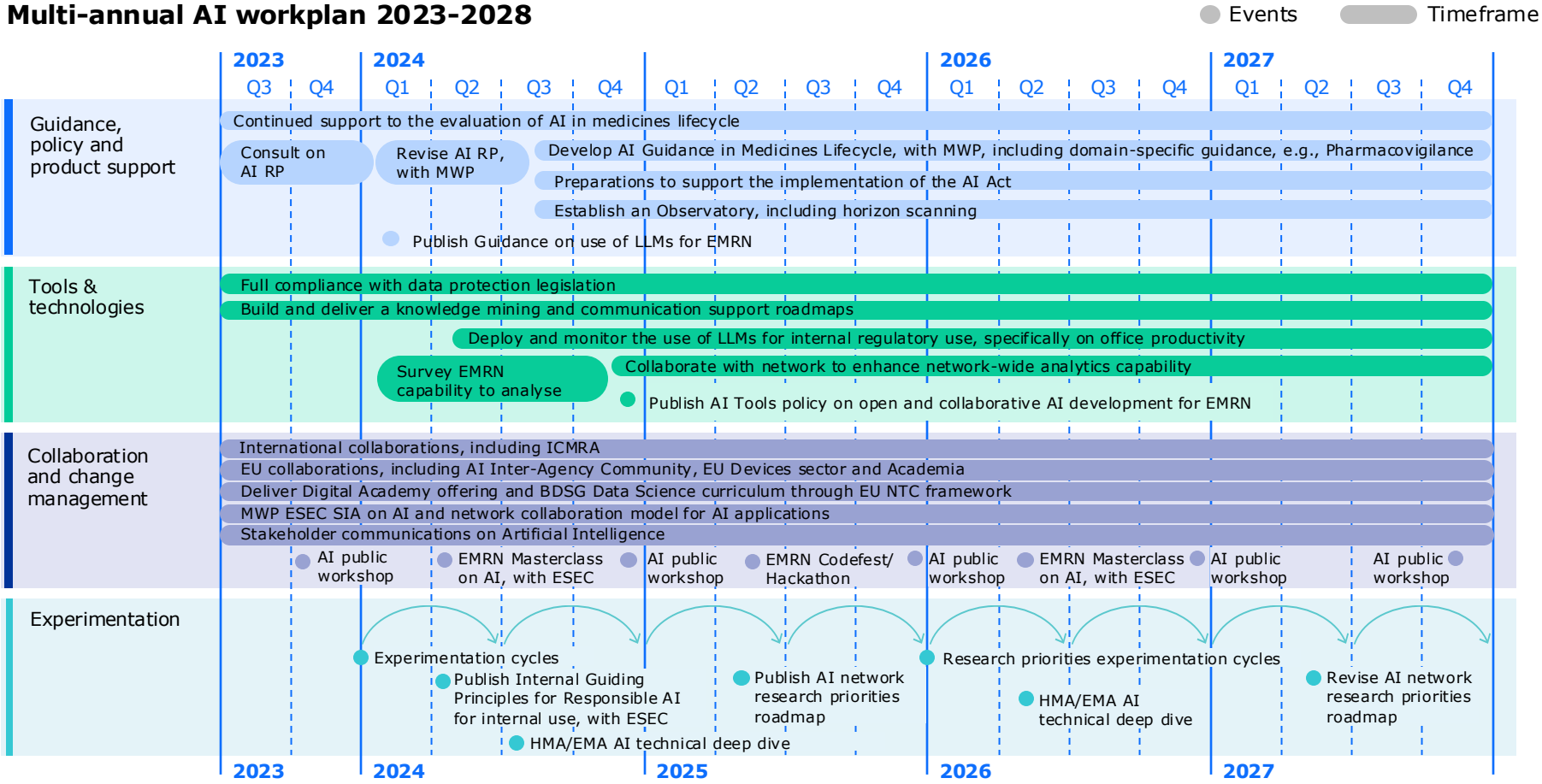
describe the broken twig syndrome

The "broken twig syndrome" is a metaphor often used in psychology and counseling to describe a phenomenon where individuals become overly focused on small problems or setbacks, to the point where they exaggerate their significance and let them overshadow everything else. Just like how a twig breaking in a forest might capture someone's attention disproportionately, this syndrome suggests that people can become fixated on minor issues, ignoring the bigger picture and positive aspects of their lives. It's often associated with pessimism, anxiety, and a tendency to catastrophize. Therapeutic interventions for this syndrome typically involve helping individuals gain perspective, develop resilience, and cultivate a more balanced outlook.

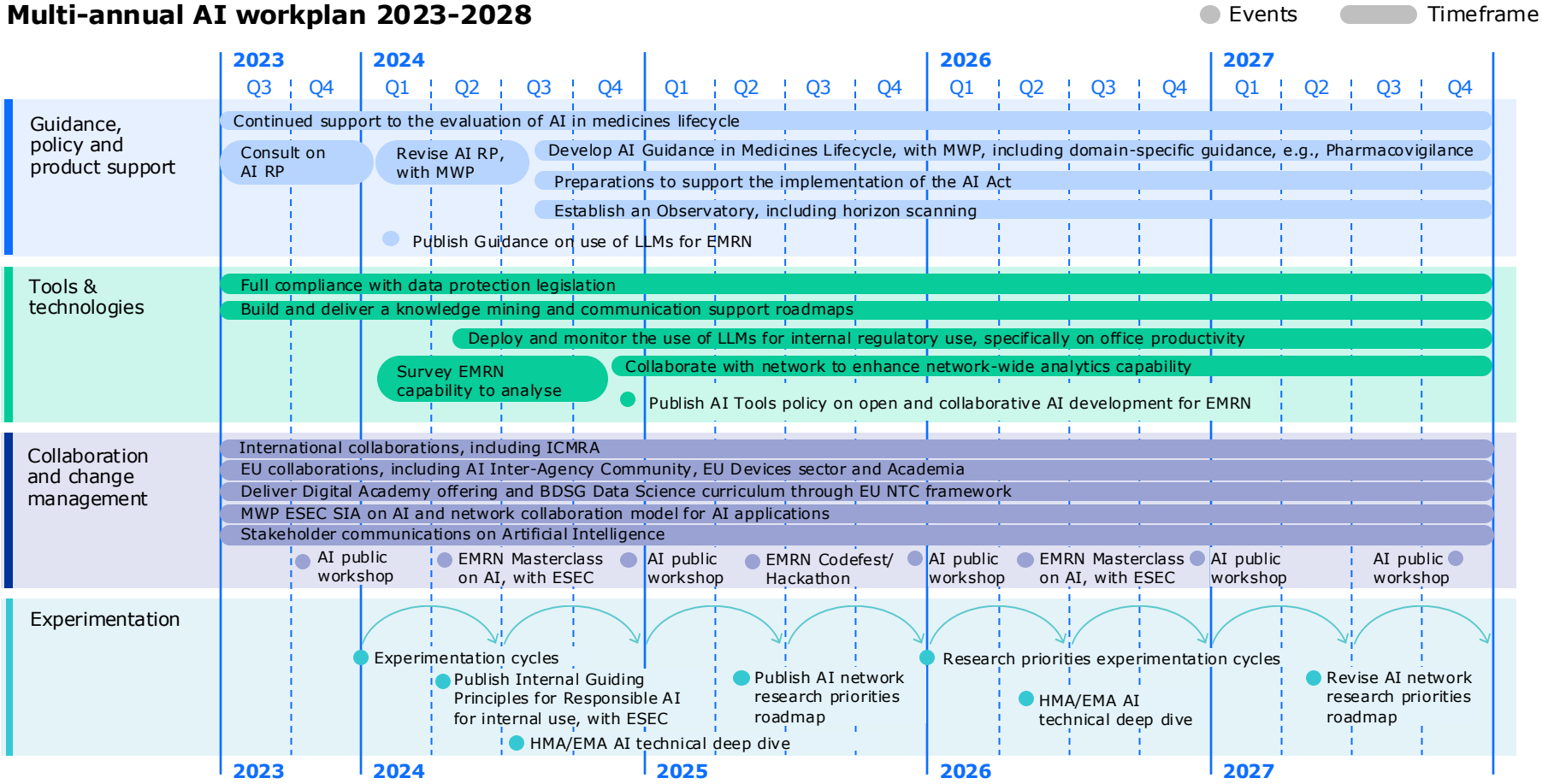
Safe and responsible use | AI not always needed



Multi-annual AI workplan 2023-2028



Multi-annual AI workplan 2023-2028



Any questions?

Further information

luis.pinheiro@ema.europa.eu

Official address Domenico Scarlattilaan 6 ● 1083 HS Amsterdam ● The Netherlands
Telephone +31 (0)88 781 6000
Send us a question Go to www.ema.europa.eu/contact

Follow us on **@EMA_News**

swissmedic 4.0

Regulator in the Loop: Trust in ML-based Systems

More-EUROPA Webinar, July 9th, 2024

Nicolas Löffler-Pérez

Schweizerisches Heilmittelinstitut
Institut suisse des produits thérapeutiques
Istituto svizzero per gli agenti terapeutici
Swiss Agency for Therapeutic Products

Hallerstrasse 7, 3012 Bern, Schweiz
www.swissmedic.ch

swissmedic 4.0

7

2020 - 2025

Swissmedic 4.0 is an innovation lab, an experimental field for change and innovation. The aim of the initiative is to promote interdisciplinary work and to design new digital business models.

“A modern regulatory agency office with a diverse team of professionals using AI technology to assess pharmaceutical products. The scene includes scientists and analysts working on computers displaying complex AI algorithms, molecular structures, and data visualizations. A large screen on the wall shows a real-time dashboard of pharmaceutical assessments. The office is well-lit with a high-tech, organized environment, featuring shelves with pharmaceutical samples and documents. The overall atmosphere is one of efficiency and cutting-edge technology.”

Automated Data Analysis:

AI can quickly and accurately analyze large volumes of data, leading to faster decision-making processes.

Risk Assessment:

AI models can assess risks more accurately and quickly by analyzing complex data and identifying potential problems early.

Potential of AI in the Regulatory Field

Monitoring and Compliance:

AI systems can continuously monitor and ensure compliance with regulations by detecting deviations in real-time.

Detection of Illegal Activities:

By analyzing online content, AI models can identify suspicious activities and products that human controllers might overlook.

Challenges in Implementing ML Systems

The introduction of ML-based tools into regulatory processes presents a significant challenge.

Contradicting principles:

These models are often perceived as "black boxes," which contradicts the principle of traceability in regulatory environments.

Competencies:

Regulators must understand not only the processes and data but also the functionality of ML models.

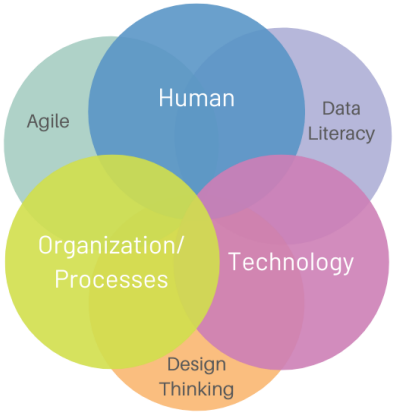
Building trust in these systems is therefore difficult.

Why Trust is Crucial

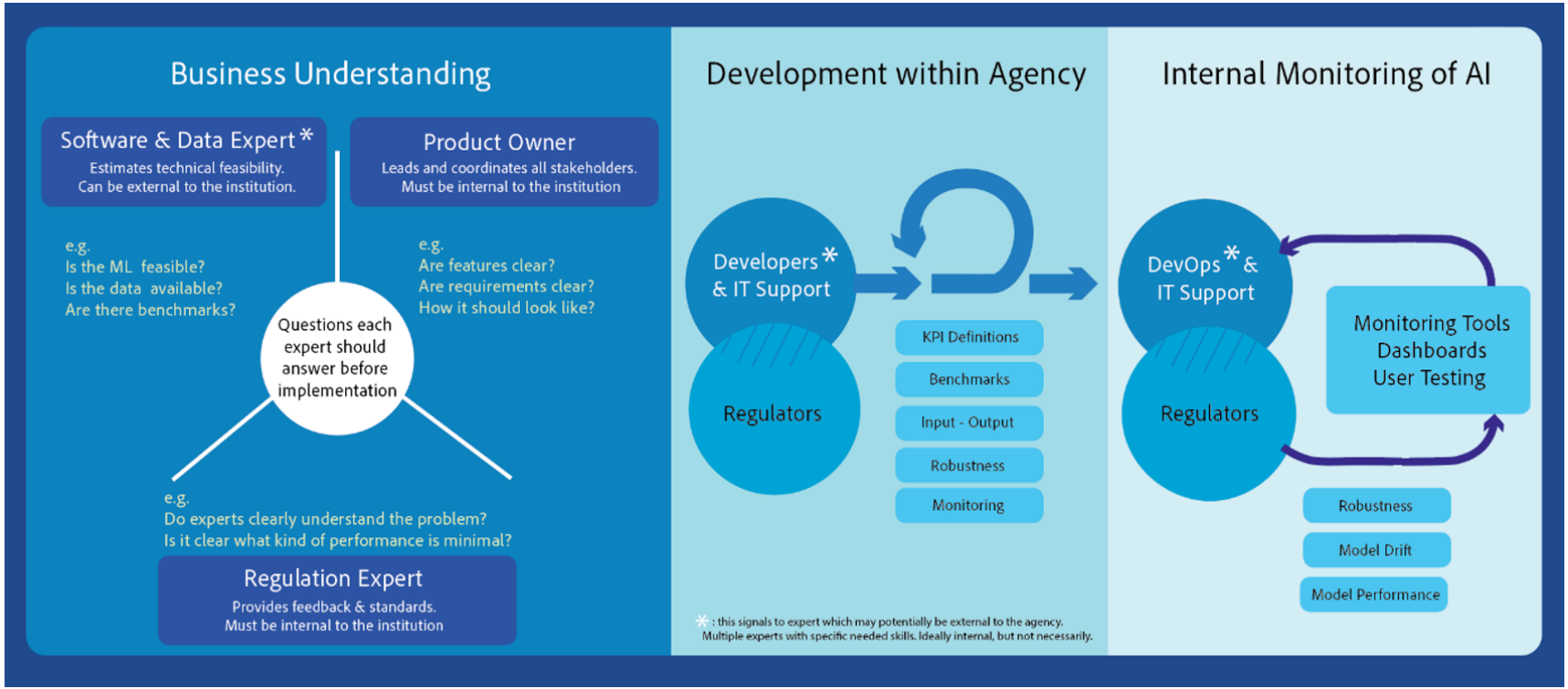
- Regulatory agencies operate in an environment where **errors cannot be tolerated**.
- Trust is crucial because regulators need to **maintain complete control and traceability** of decision-making processes.
- **A system that regulators do not trust** will not be used effectively and can **lead to uncertainties** and potentially serious misjudgments.
- Trust is built through transparency, traceability, and the continuous involvement of users in the development process.

- Comparison with human working relationships: **Trust through openness and communication**.

Our Framework



- We have developed a framework that supports the introduction and operation of ML-based systems in regulatory environments
- The central theme of our approach is trust. Trust in the technology and its results is essential, comparable to the trust that is indispensable in human working relationships.
- Our framework is based on two main components:
  - Technical Backbone: Performance metrics, transparency requirements, data management
  - Iterative “Show-and-Tell”: Agile methods, involvement of regulators, transparency, and trust


Business Understanding
Software & Data Expert *
Estimates technical feasibility.
Can be external to the institution.
Product Owner
Leads and coordinates all stakeholders.
Must be internal to the institution
e.g.
Is the ML feasible?
Is the data available?
Are there benchmarks?
e.g.
Are features clear?
Are requirements clear?
How it should look like?
Questions each expert should answer before implementation
e.g.
Do experts clearly understand the problem?
Is it clear what kind of performance is minimal?
Regulation Expert
Provides feedback & standards.
Must be internal to the institution
Development within Agency
Developers * & IT Support
Regulators
KPI Definitions
Benchmarks
Input - Output
Robustness
Monitoring
* : this signals to expert which may potentially be external to the agency.
Multiple experts with specific needed skills. Ideally internal, but not necessarily.
Internal Monitoring of AI
DevOps * & IT Support
Regulators
Monitoring Tools
Dashboards
User Testing
Robustness
Model Drift
Model Performance

Use Case 1: /TRICIA

The goal of TRICIA is the risk-based triage of medical device incidents.

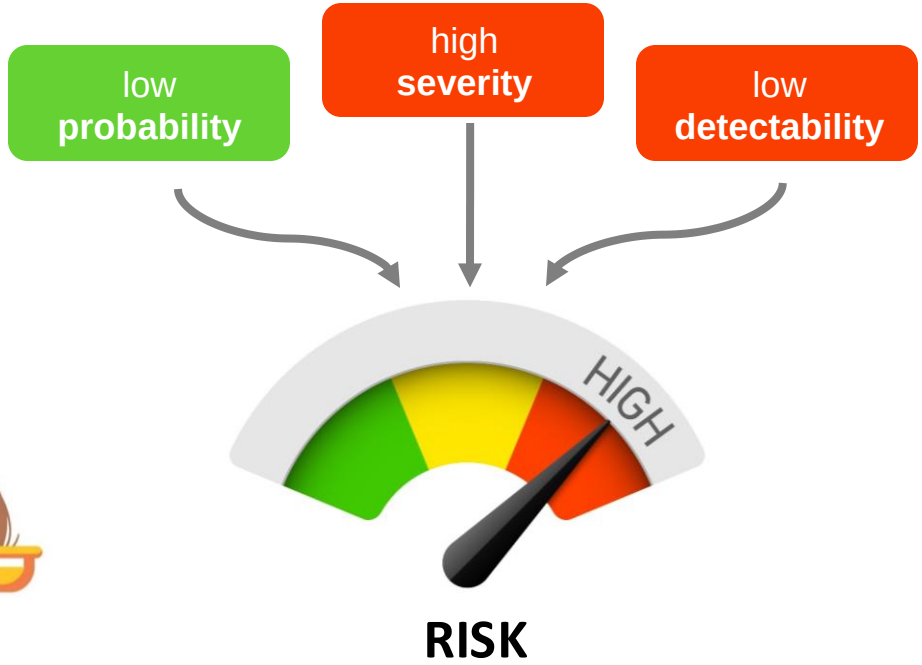


Image source: shutterstock

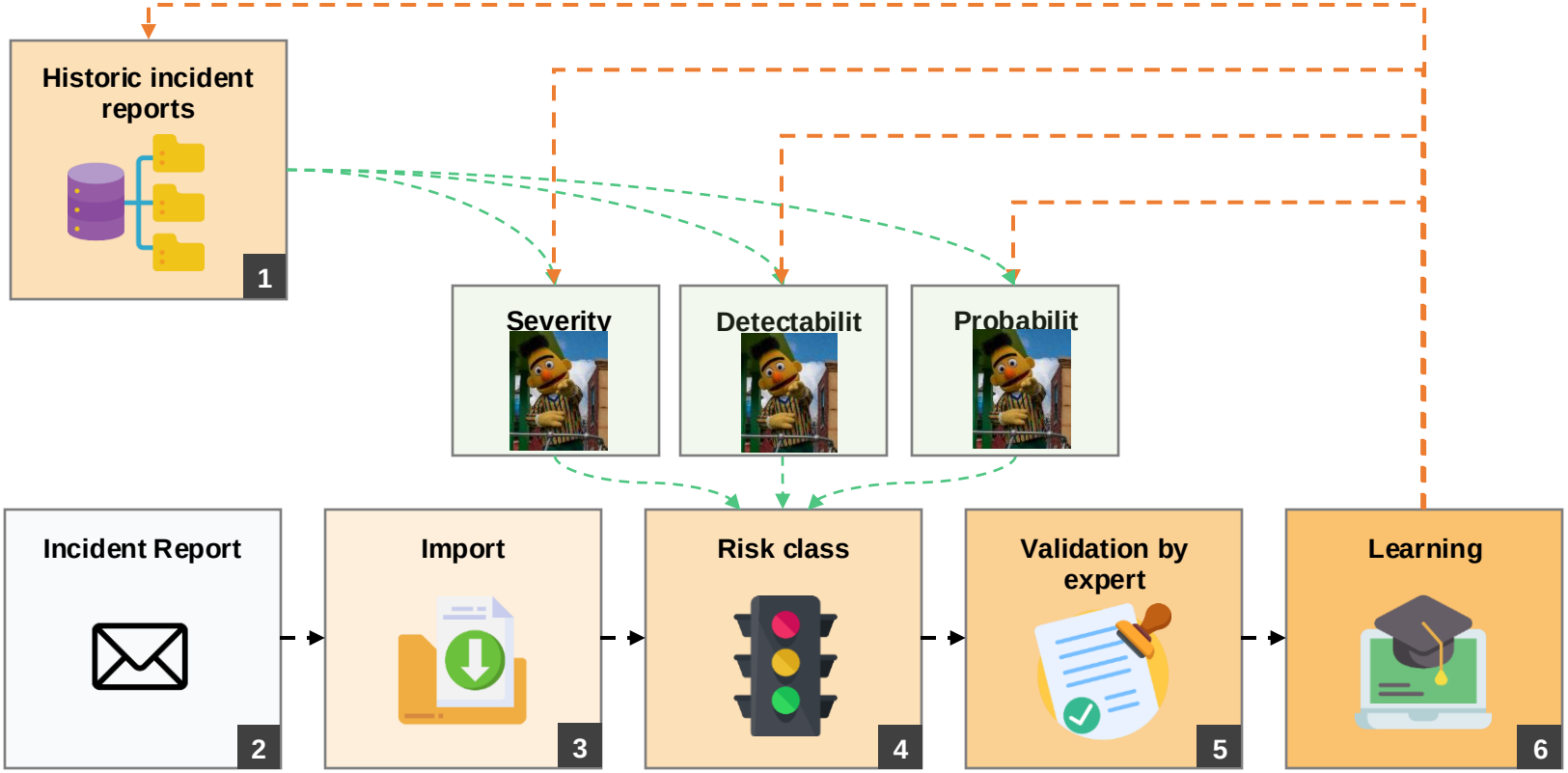


Image source: shutterstock

Use Case 1: /TRICIA

The goal of TRICIA is the risk-based triage of medical device incidents.

- In the beginning, we had **~10,000 entries available as the gold standard**, collected over four years.
- We performed multiple data analysis approaches.
- The ML model was trained to assess incidents based on risk criteria.
- Two main KPIs were defined: **the detection rate of high-risk cases and the F1-score multi-class evaluation**.

- KPIs are challenged and replaced for more meaningful ones.

Version 1.0

- Through regular feedback from regulators and the adaptation of the model to new data and requirements, the system was continuously optimized.
- Through experimentation, experts could discover inconsistencies from manual work and the advantages of ML.
- Monitoring was key for trust

Version 2.0

- Experts remain in the driver seat, as per defining which features on the application remain crucial for performance.

Use Case 2: /NightCrawler

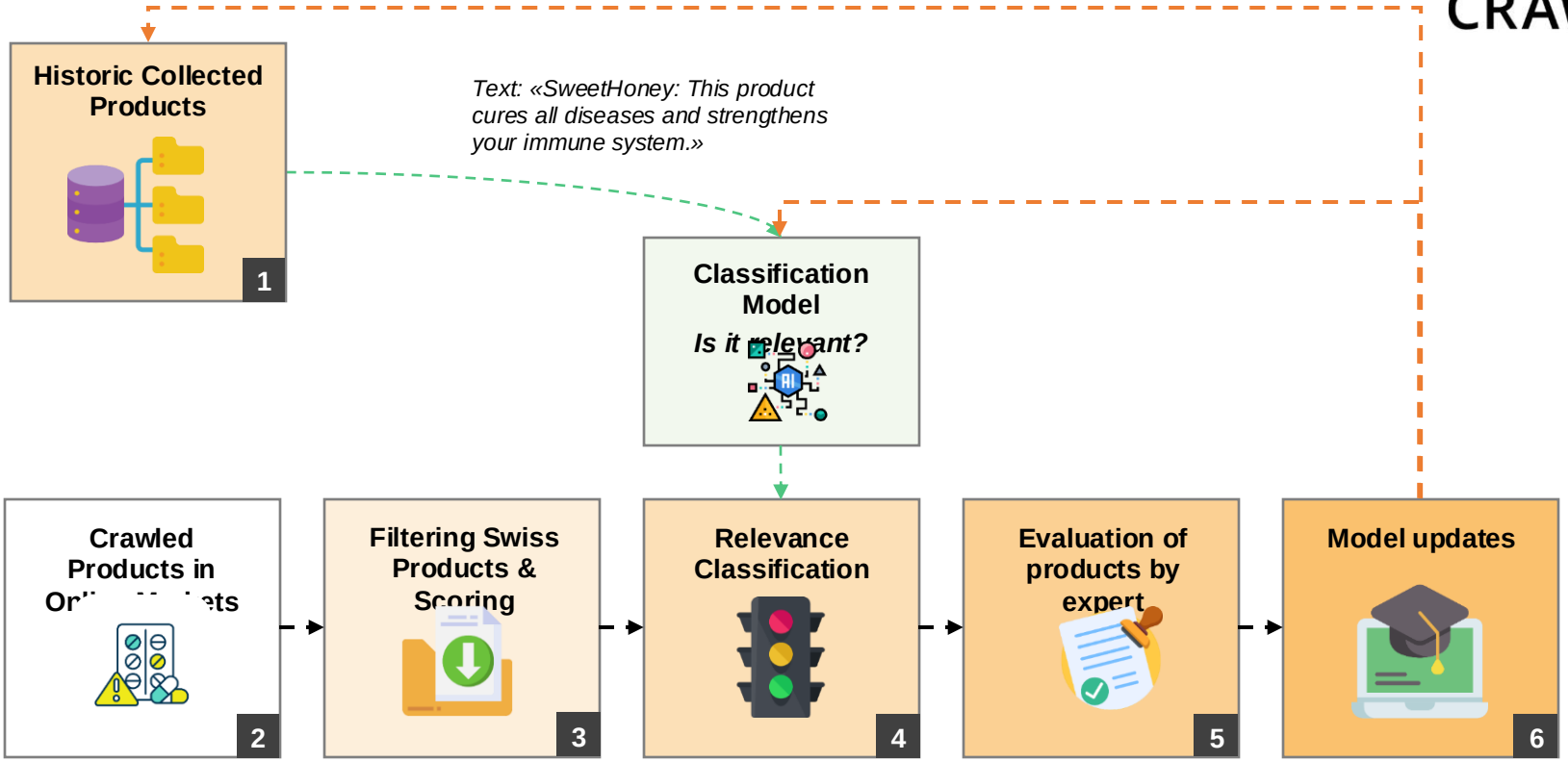
The goal of NightCrawler is to automate the search for illegal medical products on the internet.

swissmedic

Image source: shutterstock

How does it work?

/MEDI CRAWL



Use Case 2: /NightCrawler

The goal of NightCrawler is to automate the search for illegal medical products on the internet.

- In the initial phase, no suitable data was available. Manually labelled data was collected. Which was a time-consuming process.
- In Version 1.0 und 2.0, the model was trained to identify and classify suspicious URLs. We placed great emphasis on the accuracy and relevance of the results.
- Over time, we collected over 50,000 manually labeled cases. These served as the basis for developing and training the ML model.
- Metrics and Benchmarks are standardized and planned for the longterm

Version 1.0	Version 2.0	Version 3.0	Version 4.0
• Experts on board since initial PoC	• Monitoring of performance and direct communication with Stakeholders allows for trust to be built on the system	• Through continuous testing and feedback from regulators, the model was continuously improved. A benchmark was established to ensure that the system found at least as many relevant URLs as human inspectors.	• Direct access to dashboard is planned with regular testing of performance.

Results

NightCrawler

- The system significantly improved the efficiency of inspectors by searching large amounts of data and filtering out relevant cases.
- NightCrawler achieved an 80% hit rate in identifying relevant URLs
- NightCrawler demonstrates how an ML tool can be developed to meet the regulators' requirements through an agile and iterative development process.

TRICIA

- The system was able to standardize and accelerate the triage process, reducing manual workload.
- TRICIA achieved high accuracy in risk assessment and improved the consistency of assessments between different experts.

Trust as the Key to Success

Implementing ML systems in regulatory environments requires a careful and transparent approach.

Trust is the key to success in implementing ML systems in regulatory environments.

By building and strengthening trust through transparency, traceability, and continuous involvement of regulators, we can effectively utilize ML systems to fulfill regulatory tasks.

Trust in technology is as important as trust in human colleagues and is built through transparency and continuous communication.

An AI-powered tool for identification, and pre-assessment, and selection of registries to support regulatory and HTA decision processes

Webinar: AI and ML approaches to identify and appraise registries and relevant data

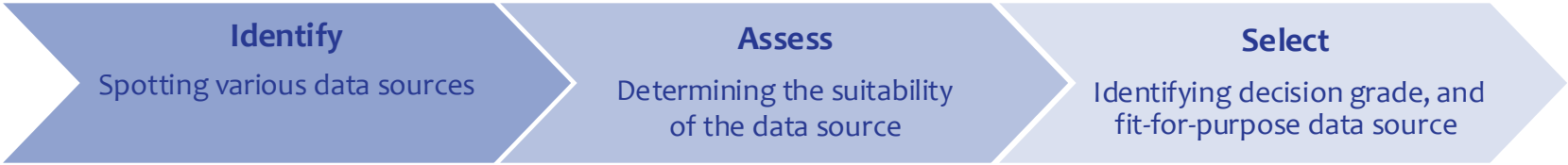
9-July-2024

Agenda

1. **The Why:** Conclusions from the needs assessment
2. **The What:** An AI-powered solution for the identification, pre-assessment, and selection of registries
3. **The How:** An AI solution enabling
   - Centralisation of different sources and types of data
   - Identification of Real-World Data through ranking and hybrid search approaches
   - Selection and pre-assessment using large language models
4. **Q&A/Feedback**

MORE EUROPA

Current needs cover: Identification, assessment & selection of fit registries



Process of registry identification:

1. Ensure the **study design** is in **adequacy** with the research question
2. Ensure the registry is **relevant and reliable**
3. Check **the ethical/legal** and practical aspects (e.g., privacy concerns, access, budget)
4. Multiple resources to search for adequate registries based different criteria such as: geographical locations, pathologies, type of studies, outcomes, population characteristics… etc

State of the Art on initiatives and tools to identify, select, & assess registries

Scoping Review 1

Objective: Detemine the existing frameworks or initiatives to identify, select, and/or assess fit-for-purpose data

Library: MEDLINE, Web of Science and Scopus

10 frameworks compliant with the FAIR principles:
RWD-Cockpit, Maturity, ATRAcTR: Authentic Transparent Relevant Accurate Track-Record, MVET principles: Meaningful – Valid - Expedited and Transparent evidence, PICOT: Population, Intervention, Comparator, Outcomes, Timing and Setting, REQuEST: Registry Evaluation & Quality Standards Tool, SPACE: Structured Pre-approval and Post approval Comparative study design, SPIFD/SPIFD2: Structured Process to Identify Fit-For-Purpose Data, SURF: Screening Tool to evaluate whether Using RWD to support an effectiveness claim in an FDA application has Regulatory Feasibility

Scoping Review 2

Objective: Look into the existing machine-learning (ML) based tools to identify or select registry using natural language processing (NLP) approaches

Library: MEDLINE and Google Scholar

✓ **63** identified tools designed to assist in the early stages of a systematic review

➢ **No automated tool t**hat perform identification, selection and assessment of registry data

Stakeholder interviews & surveys

Objective: Capture current practices and unmet needs to identify, select and assess registries
Methods: 12 stakeholders (regulators, academicians, researchers, industrials) participated in a KAP Survey followed by a 1h interview with open-ended questions

✓ **2 frameworks identified:**
- ➢ The Duke Margolis Center for Health Policy framework
- ➢ The Health Data Research United Kingdom 'data utility' framework

✓ **Main gaps in:**
- ➢ Discoverability & Accessibility
- ➢ Relevancy & Reliability

The Why: Existing solutions do not address some unmet needs identified

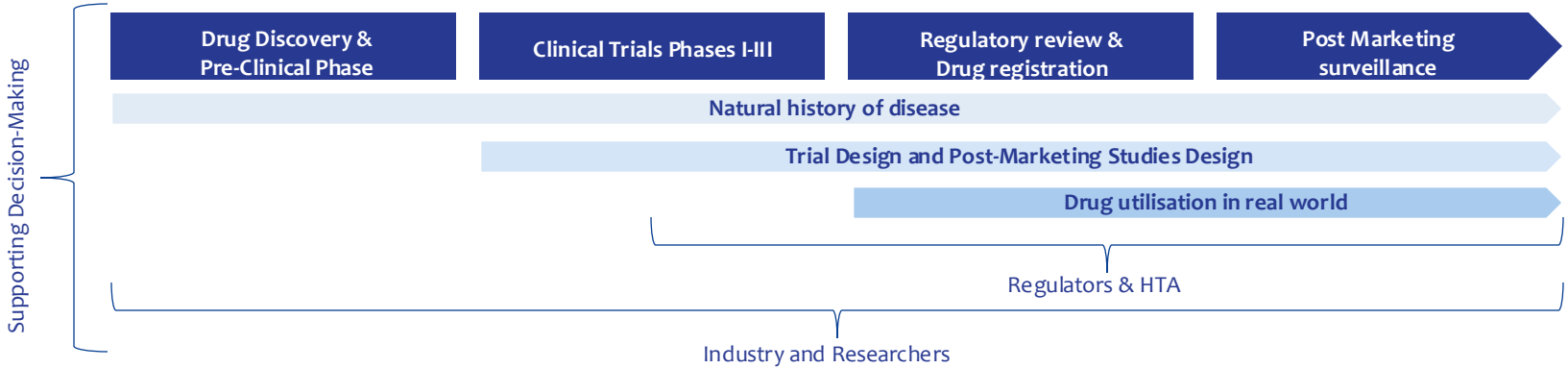
Type	Catalogue/Resources Name	Authentication needed	Coverage
Catalogues including registries	HMA-EMA Catalogues of real-world data sources and studies	Not needed	+
	European Health Data & Evidence Network (EHDEN) portal	Needed	+
Disease based catalogues	European Platform on Rare Disease Registration (ERDRI.dor)	Needed	-
	Orphanet: portal for rare diseases and orphan drugs with registries and biobanks	Not needed	+
	Multiple Sclerosis Data Alliance (MSDA) Catalogue	Needed	+
Catalogues or resources including observational studies derived from registries	ClinicalTrials.gov portal	Not needed	++
	PubMed – literature database of biomedical and life sciences	Not needed	++
	Semantic Scholar - literature database	Not needed	++
Private Initiatives	IQVIA Health Data Catalog	Needed	++

What: One solution addressing multiple needs for multiple user profiles & decision processes

Mission: Enhancing the identification, selection, and pre-assessment of registries

Vision: An open source, AI-powered, single-point-of-access tool to rapidly identify and pre-assess registries for regulatory and HTA purposes

Drug Development Lifecycle x Potential End Users



Proposed added value:

- ✓**Accessible and open-source tool:** be aligned with FAIR (Findable, Accessible, Interoperable, Reusable) principles
- ✓**One-stop shop web-based platform:** collection and analysis of information from diverse sources
- ✓**AI-powered tool:** boost time saving and actionability of findings
- ✓**User-centered approach:** tool developed with and for users

The **what**: An AI-driven tool to centralise, identify, & pre-assess registries

Centralise
Provide a single-point-of-access with up-to-date information

Identify
Provide results aligned with user search

Pre-assess
Provide useful information for the user

Scattering of information

Relevancy of information

Accessing up-to-date information

Automated data extraction

Hybrid search

Language model

3

The how: A single-point-of-access with up-to-date information

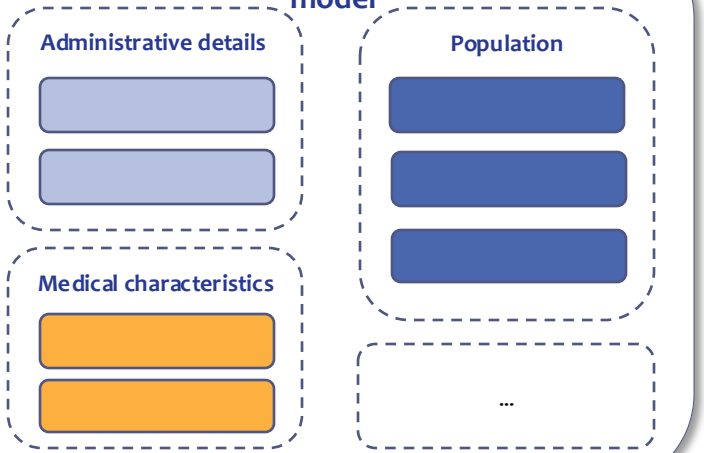
Scattering of information

- Few institutional catalogues
- Limited number of listed data sources
- Need to complement the search with observational studies or scientific literature
- No one-stop-shop platform

△ Real-world data sources
○ Observational studies
□ Publications

Weekly extraction

Automated data extraction



This data centralisation process is an entry point towards identification and pre-assessment

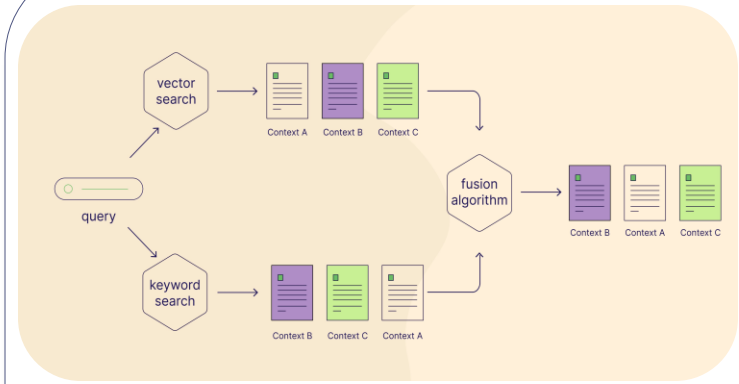
3

The how: Augment standard keyword search with AI

Relevance of information

- Capture **context** and **intent** behind the user query (e.g., synonym, semantically close word)
- Discover **content not explicitly searched for** but still valuable
- Save time by **prioritising** results
- Save time by **avoiding writing long and complex** (PubMed) **query**
- Reduce **false positive** (e.g., a register / to register)

Hybrid search



Combine traditional keyword search with modern vector search to provide more accurate and relevant results

A hybrid text search combined with filters to enhance the identification process

The how: Augment standard keyword search with AI

Hybrid search

Keyword search through BM25 algorithm (e.g., PubMed)

Keyword gets a score for each document that is

proportional to:

- The **keyword frequency in the document**

and inversely proportional to

- The **length of the document**
- The **keyword frequency in all documents in database**

A limiting example:

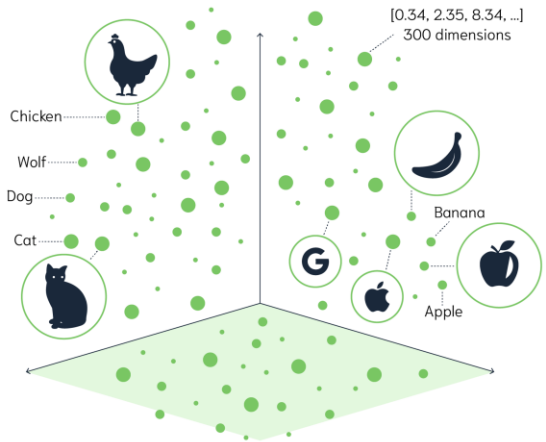
Data

- Procedure: Cardiac Catheterisation
- Description: … used to diagnose and treat certain cardiovascular conditions …

Query

- "Procedure for heart disease"

Vector search in high dimensional space (e.g., recommendation platforms)



Close representations have higher scores (e.g., heart, blood, cardiovascular, cardiac)

3

The More-EUROPA tool

Quick live demo 1

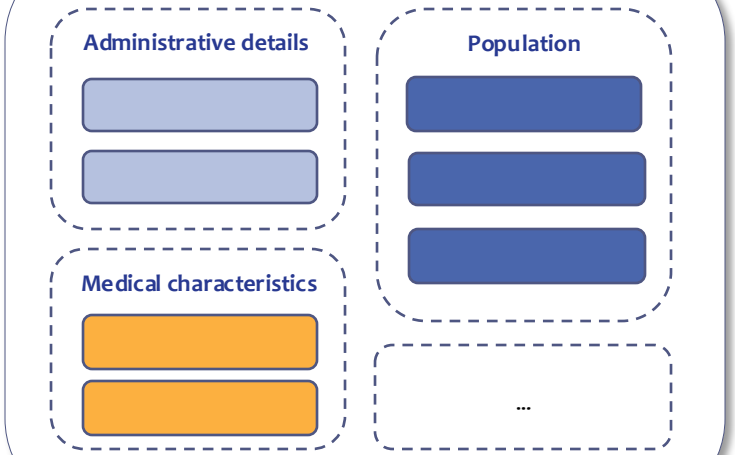
3

The how: Common metadata model enables quick & robust identification & pre-assessment

Accessing up-to-date information

- Metadata provides a quick **overview of the registry** during the pre-assessment process
- The common metadata model makes it possible to **compare registries quickly**:
  - The same information is available for all data of the same type
  - Much of it, is common to all 3 types of data
- Metadata makes it possible **to refine the search** filtering on it

Common metadata model



3

The **how**: LLMs a powerful tool to augment & accelerate key data extraction & processing

Accessing up-to-date information

Publications particularity:

- Mainly **unstructured** data (abstract) comparing with other types of data
- More difficult to refine the search: **limited amount of metadata** and only few possible filters
- More difficult to pre-assess a registry: **laborious to find information** about registries in the text

Language model

Metadata **automatically extracted from abstracts** with a **Large Language Model** (LLM):

- Facilitates registry **pre-assessment**
- Offers **new filters**

Publication's medatada

New structured metadata:

Population (size, age, sex), Intervension, Comparator, Outcome/Endpoint (PICO), setting: geographical area…..

The **how**: LLMs a powerful tool to augment & accelerate key data extraction & processing

Language model

Abstract

Effectiveness of TNF inhibitor switch in RA: results from the national Swedish register

Abstract

Background: Switching to a second tumour necrosis factor inhibitor (TNFi) after discontinuation of a first in rheumatoid arthritis (RA) is a common strategy. The reason for the switch from the first TNFi could potentially influence the response to therapy. Data on direct comparisons between TNFi after switching are limited.

Methods: The national Swedish register was used. RA patients who switched to a second TNFi (infliximab, etanercept or adalimumab) after failure of a TNFi as first-ever biologic were identified. Effectiveness of treatment was compared across the three drugs according to the first TNFi used, the reason for discontinuing and the drug survival. Drug survival across TNFi used as second biologic was compared.

Results: Half of all patients starting infliximab, adalimumab or etanercept during the period 2005-2012 discontinued treatment for various reasons. Of these patients, a third switched within 2 months to a second TNFi (infliximab, etanercept or adalimumab). Around 35% of all patients achieved low disease activity or remission at 6 months. Regarding the switching strategy, best results were observed among patients who switched from infliximab to etanercept because of (secondary) inefficacy. Etanercept as second TNFi was associated with longer drug survival compared with infliximab. [...]

LLM (type of NLP models)

Metadata

- Registry name: National Swedish register
- Population: Patients with rheumatoid arthritis (RA) who switched to a second tumour necrosis factor inhibitor after discontinuation of a first TNFi.
- Condition: Rheumatoid arthritis
- Sex: Not specified
- Age group: Not specified
- Size: Not specified
- Intervention: Not applicable
- Comparator: Infliximab, etanercept, or adalimumab as second TNFi
- Outcome/Endpoint: Drug survival, low disease activity or remission at 6 months
- Country: Sweden
- Follow-up: Not specified

Warning: information extracted automatically from an LLM may be incomplete or not accurate.

The More-EUROPA tool

Quick live demo 2

An AI-powered tool to centralise, identify, & pre-assess registries

Centralise
Provide a single-point-of-access with up-to-date information

Identify
Provide results aligned with user search

Pre-assess
Provide useful information for the user

Scattering of information

Automated data extraction

One stop-shop: Boost time saving and avoid multiple searches

Relevancy of information

Hybrid search

Ameliorate the search and prioritise relevant results

Accessing up-to-date information

Language model

Enhance information retrieval and actionability

Thank you

MORE EUROPA

Q&A
Feedback

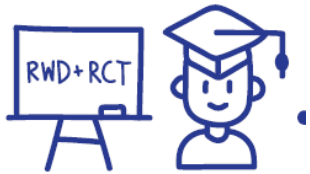
Q&A session

Moderator: Cécile Ollivier

Panelists: Peter Mol, Luis Pinheiro, Nicolás Pérez González, Billy Amzal, Ghinwa Hayek, Pascal Godbillot, Coriande Clemente

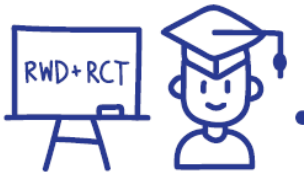




Wrap-up and conclusion

Peter Mol



Thank you!

Learn more: https://more-europa.org/

Building the future of health

MORE EUROPA

umcg