GETREAL INSTITUTE | PUTTING REAL WORLD EVIDENCE INTO PRACTICE

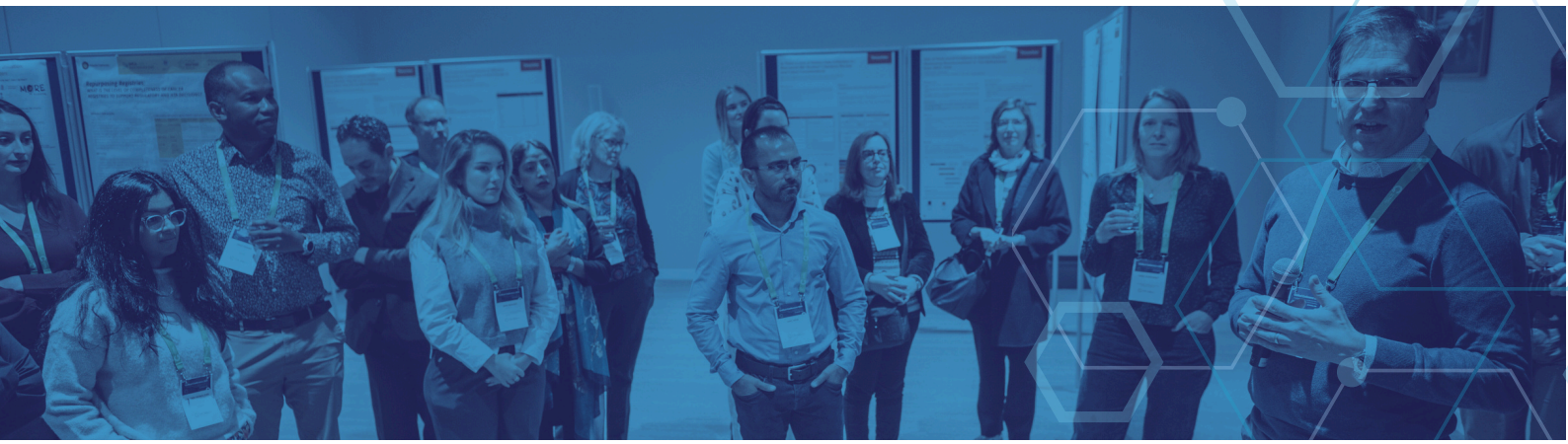
GETREAL CONFERENCE 2025

ABSTRACTS | POSTER BOOK

GetReal Conference 2025:
Selection of accepted posters

1-2 October 2025 | Utrecht, The Netherlands

CONTENT



INTRODUCTION OF **RESEARCH ABSTRACTS AND POSTERS**

As part of the 2025 conference programme, the GetReal Institute introduced a call for research abstracts for the first time, expanding opportunities to showcase the breadth of activity across the real-world evidence (RWE) community. Submissions spanned applied case studies, pilot projects, and methodological developments, reflecting the diversity of approaches used to generate and apply RWE. This initiative reinforced the Institute's commitment to advancing best practice, encouraging dialogue, and broadening engagement across the RWE ecosystem.

Fostering Cross-Sector Collaboration

The abstracts and poster track provided a structured forum for exchange between researchers, regulators, HTA bodies, patient representatives, and industry stakeholders. Contributors shared experiences and practical approaches from different perspectives, supporting the Institute's mission to strengthen cross-stakeholder understanding and reduce silos in the generation and use of RWE.

Highlighting Innovation and Best Practice

The call for abstracts highlighted innovative concepts, practical solutions, and emerging methodologies relevant to regulatory, HTA, and broader decision-making contexts. The selected contributions demonstrated how RWE continues to evolve in response to methodological and policy challenges.

Supporting Early Career Engagement

The introduction of a poster track created dedicated opportunities for early-career professionals and students to present their work, engage with senior experts, and expand their professional networks, contributing to capacity building within the RWE community.

REVIEW & SELECTION PROCESS

All submitted abstracts were reviewed and scored by the GetReal abstract committee, comprising representatives from across the Institute's membership. Abstracts were evaluated against predefined criteria, including relevance, methodological quality, innovation, and potential impact. Following this process, three abstracts were selected for oral presentation at the conference:

- Repurposing Registries: What is the level of maturity of cancer registries to deliver Real World Evidence to Support Regulatory and Health Technology Assessment decisions. Presenter: Manon Wilpshaar, Utrecht University, The Netherlands

- REALM: From Black Boxes to Transparent Pipelines. Presenter: Anshu Ankolekar, Maastricht University, The Netherlands

- The pilot of the MHRA Real-World Evidence Scientific Dialogue Programme for medicinal products. Presenter: Doaa Elkholly, MHRA, England

POSTER SESSION

A moderated poster presentation and networking session was held on Day 2. This session enabled authors to present their work, receive feedback, and engage in interactive discussion with conference participants.

POSTERS

Selection of Accepted Abstracts & Posters GetReal Conference 2025.
This publication includes only those posters for which authors granted permission.

- Low left truncation mitigation rates in post-marketing pregnancy exposure registries assessing the risk of spontaneous abortion (SAB)
- Improving the transparency and ethics of studies using real-world data
- The pilot of the MHRA Real-World Evidence Scientific Dialogue Programme for medicinal products
- Integration of real-world data in reimbursement decisions across Europe: national variability, frameworks, and emerging challenges
- Optimizing Electronic Medical Records Data Completeness with On-Premise Artificial Intelligence: A Study on Large Language Models Enhancing Medication Documentation
- Unlocking the Potential of Radiology Reports using NLP: A Real-World Data Approach to Rotator Cuff Tear Severity
- Leveraging Real-World Data and NLP for Early Identification of Fatty Liver Disease in Radiology Reports
- MORE Europa: More Effective and ethical Use of Registry data to suppOrt PAtient-centred regulatory and HTA decision making
- Using large language models for scalable extraction of real-world progression events across multiple cancer types
- European perspectives on ethical aspects around utilization of patient registries for medicines decision-making
- Repurposing Registries: What is the level of maturity of cancer registries to deliver Real World Evidence to Support Regulatory and Health Technology Assessment decisions
- REALM: From Black Boxes to Transparent Pipelines
- Characterization of Novel Longitudinal Oncology Real World Datasets in Germany and the UK
- Integrating Real-World Evidence For Advanced Therapies In Inflammatory Bowel Disease Into Re-Imbursement Submissions: Insights And Considerations From The UK
- Use of Real-World Evidence in Advice Reports Assessing Reimbursement in The Netherlands from 2023 – 2025

POSTERS

Selection of Accepted Abstracts & Posters GetReal Conference 2025.
This publication includes only those posters for which authors granted permission.

- Is There a Lack of Dietary Data Collection in Real-World IBD Studies? Literature Review and Future Considerations
- Cross-Sectional an Efficient to Prospective Cohort Design in Real World Studies (RWS)?.
- Data Reliability In Retrospective Chart Review Studies: Results And Considerations From A Novel Data Review Methodology
- Cost-Effectiveness of Radiotherapy in Uterine Serous Carcinoma (USC): A Real-World Study.
- A Real-World derived algorithm to improve the Cost-Effectiveness of External Beam Radiotherapy in Uterine Serous Carcinoma
- Comparing Randomized Clinical Trials to Real-world Studies Evaluating Effectiveness of a bDMARD in the Management of Crohn's Disease
- Comparison of Randomized Controlled Trials (RCT) with Open Label Single Cohort Real World Studies (RWS) in the evaluation of pharmacologic treatment of Non-Small Cell Lung Cancer (NSCLC)
- Utility of Real World External Control Arm (ECA) in the evaluation of pharmacologic treatment of Non-Small Cell Lung Cancer (NSCLC)
- Advancing Real-World Evidence with Causal AI: A Scalable Framework for Unlocking Actionable Insights from Real-World Data

Low-left truncation mitigation rates in post-marketing pregnancy exposure registries assessing the risk of spontaneous abortion (SAB)

ABS-01

Maria Angeles Natividad Sancho,[1] Reem Masarwa[2]
[1]Thermo Fisher Scientific, Paris, France; [2]Thermo Fisher Scientific, Montreal, Canada

Background

- The risk for spontaneous abortion (SAB) in post-marketing safety studies is often evaluated through pregnancy exposure registries, where recruitment occurs after pregnancy recognition. However, most SABs occur early in pregnancy, often before pregnant women enroll into the registry,[1] and the risk for SAB rapidly decreases across the first trimester of pregnancy (left truncation). If not accounted for during the analyses, left truncation can lead to inaccurate conclusions or flawed decision-making.[2-4]

Objective

- **To review the literature on approaches to the analysis of SAB risk in prospective observational studies, including pregnancy exposure registries.**

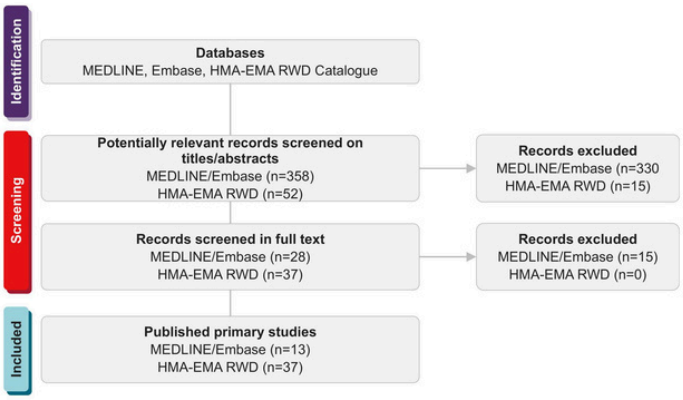
Methods

- A structured review of the literature was conducted in July 2025 in the Heads of Medicines Agencies (HMA)-European Medicines Agency (EMA) Catalogue for real-world data (RWD) sources, Embase, and MEDLINE. The search was based on explicit keywords and eligibility criteria and used a search strategy that was developed in consultation with a research librarian. The goal was to summarize the methodology used to estimate the incidence and prevalence of SABs, as well as author-reported limitations related to the potential impact on current SAB epidemiologic estimates.
- Peer-reviewed, full-text original research articles indexed in MEDLINE and Embase were included if they met the following criteria: (1) human study; (2) study designs: prospective; (3) reporting occurrence measures (prevalence, incidence, proportions); (4) reporting early pregnancy loss, spontaneous abortion, or miscarriage; (5) English-language studies; and (6) published between 1974 and July 2025 (Embase) and 1946 to July 2025 (MEDLINE). Commentaries, clinical trials, case series, case reports, and studies reporting data infertility, in vitro fertilization, or assisted reproductive techniques were excluded.
- We also reviewed data from protocols and reports from all pregnancy registry studies registered in the HMA-EMA RWD Catalogues until July 2025 that were required by regulators.
- Titles, abstracts, and full texts of the potentially relevant articles and protocols/reports were screened by a single researcher to identify those that met the eligibility criteria.
- Data on study characteristics, methodology, population, and outcomes of interest were extracted from eligible publications and protocols/reports into a bespoke spreadsheet. Limitations reported by the authors were extracted as free text and grouped by topic for the purposes of evaluation and discussion.

Results

- Our search generated 52 potentially relevant protocols from pregnancy exposure registries from the HMA-EMA Catalogue for RWD sources and 358 peer-reviewed articles from MEDLINE/Embase. After screening, we included 37 (71%) pregnancy exposure registries published between 2011 and 2024, and 13 (4%) peer-reviewed manuscripts published between 2009 and 2023 (**Figure 1**).

Figure 1: Study search and screening flow diagram



Abbreviations: EMA = European Medicines Agency; HMA = Heads of Medicines Agencies; RWD = real-world data

- Of the 37 pregnancy studies registered in the HMA-EMA Catalogue, most (78%) were required by regulatory authorities. When examining drug exposure, 35% reported SABs among pregnant women exposed to treatment for autoimmune diseases, 22% to vaccines, and 14% to antimigraine drugs. Sixty-five percent of the registries were conducted in North America and 27% in Europe; cohort sample sizes ranged from 50 to >1000 pregnancies.

Conclusions

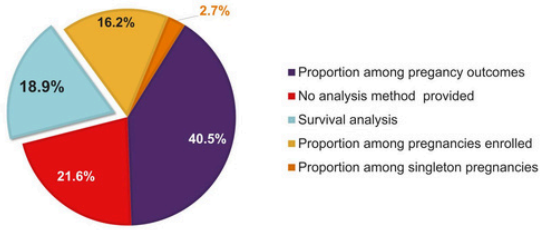
- **Left truncation, common to studies evaluating SAB rates, needs to be properly handled using survival analysis methods to avoid bias. Otherwise, when used in comparative analyses, the relative risk of SAB calculated from cohorts differentially impacted by this bias may be misleading.**
- **Only a small proportion of pregnancy registries required regulatory authorities, and an even smaller proportion of published studies that assessed SAB in the general population, used survival analysis methods to minimize left truncation bias when estimating SAB rates.**
- **A significant proportion of pregnancy registries, and fewer article reports, that did not account for left truncation in the analysis acknowledged it as a study limitation, suggesting there is a potential barrier to the use of survival analysis.**

Results (cont.)

HMA-EMA Catalogue pregnancy registries

- Among the registries that described a SAB analysis (n=29), only 24% (n=7) implemented survival analysis methods to calculate SAB rates in one sample or ≥2 sample comparisons (e.g. exposed vs unexposed cohorts) (**Figure 2**). Of these, 57% (n=4) explicitly mentioned that survival analysis methods were used to handle left truncation, as well as right-censoring when a participant was lost-to-follow-up prior to 20 weeks' gestation or when the participant was no longer at risk of the event. An additional study (14%) performed a sensitivity analysis by gestational age at enrollment.

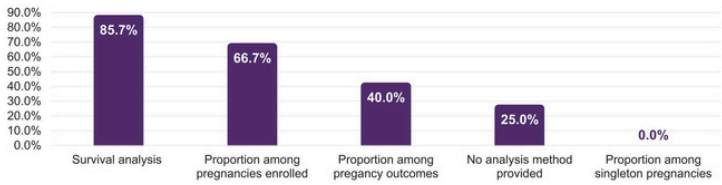
Figure 2: Calculation of SAB rates among HMA-EMA Catalogue pregnancy registries



Abbreviations: EMA = European Medicines Agency; HMA = Heads of Medicines Agencies; RWD = real-world data; SAB = spontaneous abortion

- Most studies that used survival analysis to account for left truncation to estimate SAB incidence acknowledged the bias associated with any other type of analysis. Although to a lesser extent, bias was also acknowledged among studies that did not use survival analysis; overall, half of the reports that did not account for left truncation in the analysis acknowledged it as a study limitation (**Figure 3**).

Figure 3: Proportion of HMA-EMA Catalogue pregnancy registries that acknowledged bias

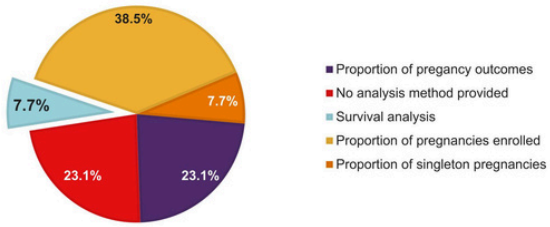


Abbreviations: EMA = European Medicines Agency; HMA = Heads of Medicines Agencies; RWD = real-world data

Embase/MEDLINE peer-reviewed articles on pregnancy studies

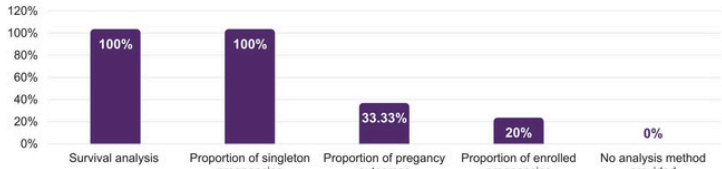
- The source data for the peer-reviewed manuscripts included national pregnancy surveillance programs (n=3), pregnancy registries (n=3), national registry databases (n=3), and hospital-based cohorts (n=4). Only 8% of reports accounted for left truncation in the analysis (**Figure 4**), and 25% of those that did not account for left truncation by survival analysis acknowledged a potential methodological weakness (**Figure 5**).

Figure 4: Calculation of SAB rates among Embase/MEDLINE peer-reviewed articles on pregnancy studies



Abbreviations: EMA = European Medicines Agency; HMA = Heads of Medicines Agencies; RWD = real-world data; SAB = spontaneous abortion

Figure 5: Proportion of Embase/MEDLINE peer-reviewed articles on pregnancy studies that acknowledged bias—by analysis method type used



Abbreviations: EMA = European Medicines Agency; HMA = Heads of Medicines Agencies; RWD = real-world data

Disclosures

MANS, and **RM** are employees of PPD™ Observational Studies, Thermo Fisher Scientific. Funding for this poster was provided by Thermo Fisher Scientific.

Acknowledgements

Editorial and graphic design support were provided by Caroline Cole and Kawthar Nakayima of Thermo Fisher Scientific.

BIH QUEST Center for Responsible Research

BIH Berlin Institute of Health @Charité

Improving the ethics and transparency of of studies using real-world data

Olmo van den Akker (ovdakker@gmail.com), Robert Thibault, John Ioannidis, Susanne Schorr, Susanne Stark, & Daniel Strech

Overview (Ethics)

Research question:
What information do patient registries provide about their ethics practices?

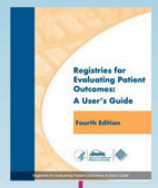
Main conclusion:
Registries provide basic ethics information, but fail to mention important details that are emphasized in existing guidelines

Overview (Transparency)

Research question:
What guidance is available for researchers to improve the transparency of studies using real-world data?

Main conclusions:
1. There is much guidance about registration and reporting but hardly any about data and code sharing
2. Challenges for improving transparency were rarely mentioned

Methodology (Ethics)



N = 51

Checklist of 26 questions categorized into 5 ethics themes:

1) Governance (3 questions)
2) Conflicts of Interest (2 questions)
3) Informed Consent (9 questions)
4) Privacy (3 questions)
5) Use-and-Access (9 questions)

→ Primarily binary coding ("Do they provide info about X?")

Methodology (Transparency)

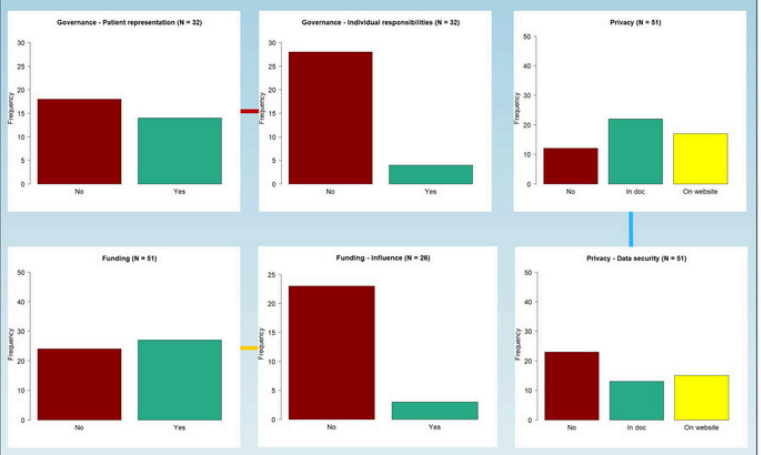
Three pillars of transparency:
- Registration
- Reporting (Methods + Results)
- Sharing (Data + Code)

Types of guidance:
- Calls
- Justifications
- Recommendations

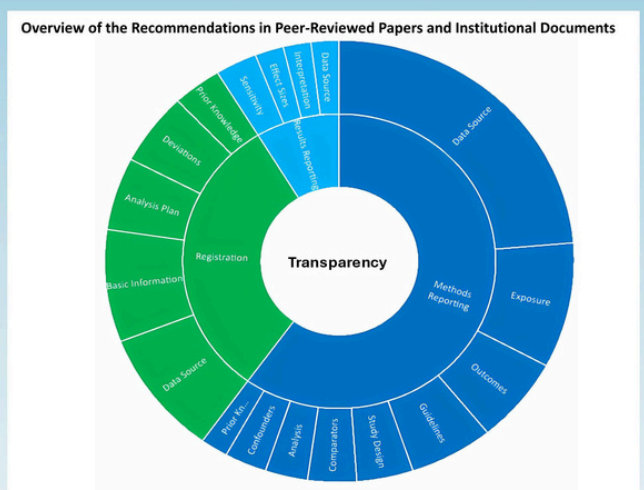
(identified in peer-reviewed papers and institutional documents)

Analysis of guidance:
- Quantitative (how many pieces of guidance?)
- Qualitative (what was the guidance?)

Selected Results (Ethics)



Selected Results (Transparency)



MORE EUROPA

MEDIZIN INFORMATIK INITIATIVE

www.bihealth.org

The pilot of the MHRA Real-World Evidence Scientific Dialogue Programme for medicinal products.

Medicines & Healthcare products Regulatory Agency

R.E. Ghosh[1], D. Elkholly [1], A. Hunt[1]

[1]Medicines and Healthcare products Regulatory Agency, Canary Wharf, London E14 4PU (MHRA)

INTRODUCTION

There is ambiguity among stakeholders surrounding the use and acceptability of Real-World Evidence (RWE) in both pre-authorisation and post-authorisation settings. Therefore, in 2025 the Medicines and Healthcare products Regulatory Agency (MHRA) piloted a RWE Scientific Dialogue Programme for medicinal products, a key deliverable of the MHRA Data Strategy (1). This was done through an expression of interest (EOI) call to assess the extent and the nature of demand for RWE regulatory dialogue.

Aims of the programme:

1. To clearly set out the MHRA's expectations for RWE methodologies for evidence generation, and to produce specific use cases.
2. To enable commercially sensitive discussions between Sponsors/MAHs and the MHRA, with a specific strategic focus on RWE.
3. To increase clarity of regulatory and Health Technology Agency's (HTA) expectations, represented by the National Institute for Health and Care Excellence (NICE), for data, analytical methodologies, and endpoints used to generate RWE.
4. To generate shared learning which can be disseminated to the broader ecosystem.

References : (1) Medicines and Healthcare products Regulatory Agency (2024). *MHRA Data Strategy 2024 -2027*. https://www.gov.uk/government/publications/mhra-data-strategy-2024-2027

METHODS

- An open call for expressions of interest (EOI) was published in the agency's web page.
- Applications were selected according to predefined priority criteria and covered a variety of use cases.

Features of the RWE SDP pilot:

Meetings

- Four closed-door, virtual meetings
- June and July
- Discuss commercially sensitive topics with strategic focus on RWE
- Enhancing existing scientific advice services.

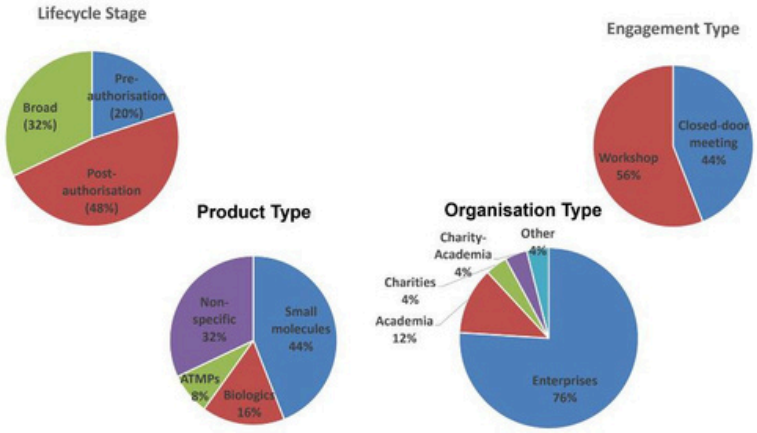
Workshop

- Hosted jointly with NICE
- Mid-July
- Discuss RWE topics relevant to medicinal products, following a review of industry expressions of interest
- For the development of collaborative resources to be disseminated to broader ecosystem

RESULTS

EOIs thematic analysis:

- The programme received 25 EOI applications.
- There was a preference for the joint workshop with NICE (56%)
- Half the EOIs were seeking dialogue at the post-authorisation stage (48%) with the remainder seeking pre-authorisation (20%) or mixed pre/post authorisation discussion (32%).



- Four closed-door meetings were held with a variety of RWE topics, disease areas and lifecycle stages.
- One workshop was convened jointly with NICE titled "using Real-World Evidence to improve therapy development for Rare Diseases: rare kidney disease as a use case."
- The workshop discussed the challenges and opportunities of using RWD rare disease registry data and surrogate endpoints for regulatory and health technology assessments.

Topics

- Data transportability
- The validity of wearable devices data
- RWD endpoints
- External control arms
- Pragmatic clinical trial design

Disease Areas

- Cardiometabolic diseases,
- Pain relief
- Large B cell lymphoma
- Opioid-induced hyperalgesia

Lifecycle stages

- Pivotal registrational trials,
- Label expansions
- Post-authorisation safety and effectiveness studies.

CONCLUSIONS

- The Real-World Evidence (RWE) Scientific Dialogue Programme pilot highlighted a strong demand for a structured regulatory engagement with the MHRA on RWE.
- The applications reflected diverse interest in using RWE to support regulatory and HTA decision-making across a wide range of medicinal product types, therapeutic areas and product lifecycle stages.
- There was a strong demand for clearer processes to qualify endpoints, methodologies and data sources, with alignment between MHRA and NICE.
- The four closed-door meetings in the RWE SDP pilot revealed diverse regulatory challenges and opportunities for RWE integration.
- The workshop underlined the interest for structured, themed engagement and there was clear demand for longer-term engagement with the MHRA.
- Applicants showed overall satisfaction with the programme as well as how the meetings and the workshop were organised. The majority of them found the programme helpful for their work.
- Insights gained from the piloted programme will be used to continue to develop MHRA Real-World Data and methodology guidance.
- A reflection paper on the workshop discussions will be published collaboratively between the MHRA, NICE and stakeholders to generate shared learning that can be disseminated to the broader eco-system. The MHRA will also develop RWE use cases from the pilot.
- Future iterations of the programme have been explored and in the planning phase.

ACKNOWLEDGEMENTS

We would like to acknowledge all the piloted programme applicants, collaborators from NICE and experts from the MHRA

Integration of real-world data in reimbursement decisions across Europe: national variability, frameworks, and emerging challenges

Sophie Boukouvalas, Jason Lucas
Avalere Health

Avalere Health™
EVERY PATIENT POSSIBLE.

Introduction

European and national initiatives increasingly promote the use of real-world data (RWD) in health technology assessments (HTAs) and policy-making, reflecting its growing importance for understanding real-world outcomes and guiding healthcare decisions. However, significant challenges remain due to diverse national frameworks, legislative environments, and data platforms across Europe.

Objectives

This research examines RWD integration in decision-making within Germany, France, Italy, Spain, and the UK by analyzing relevant institutions, legislation, and data infrastructure.

Methods

Official websites of national health authorities, regulatory agencies and relevant government bodies were reviewed (Table 1) for guidelines on RWD utilization, legislative frameworks governing RWD utilization, and descriptions of supporting data platforms.

Table 1: **Institutional, regulatory and data governance bodies reviewed**

	Federal Ministry of Health (BMG), and Federal Ministry for Research, Technology and Space (BMFTR)
	Spanish Agency of Medicines and Medical Devices (AEMPS), and Ministry of Health (MS)
	Italian Medicines Agency (AIFA)
	French National Authority for Health (HAS), and French Republic (RF)
	National Institute for Health and Care Excellence (NICE), National Health System (NHS, UK), and GOV.UK

BMFTR, Bundesministerium für Forschung, Technologie und Raumfahrt; BMG, Bundesministerium für Gesundheit; MS, Ministero della Salute; RF, République Française

Results

European countries show significant variability in integrating RWD into decision-making due to differing national approaches and acceptance levels

Table 2: **National approaches to RWD use in HTA decision-making**

	The G-BA remains ***cautious about RWE*** as a primary source of evidence and continuing to favor RCTs.
	The ***Valtermed system*** collects RWD from all the autonomous Spanish healthcare regions, providing a standardized platform to inform decisions on the therapeutic value of medicines with significant clinical and economic impact.[1]
	Recent steps includes new AIFA ***guidelines*** on observational studies, reflecting growing recognition of the integrative role of RWE.[2]
	HAS has become ***increasingly receptive*** to RWE, particularly in post-registration (post-marketing) studies to address uncertainties in efficacy, safety, or real-world use after initial registration.[3]
	NICE adopts a ***progressive approach*** by actively incorporating RWE when it complements trial data or provides unique insights.

Unlike the US's centralized approach, EU members states navigate complex national frameworks. Each country balances leveraging real-world insights with maintaining rigorous standards for evidence quality and applicability

Table 3: **National frameworks for RWD use in European HTA**

	The G-BA remains ***conservative*** in its approach, but increasingly acknowledges the value of RWD, especially where RCTs are not feasible or sufficient. The ***Health Data Usage Act*** aims to support research by enabling linked data access via the ***Health Data Lab***.[4,5]
	Efforts are underway to standardize RWD use in HTA submissions despite regional variability, with initiatives such as the ***AEMPS Strategic Plan 2023–2026*** driving progress.[6]
	AIFA addressed the lack of a standalone RWE framework by publishing ***formal guidelines on observational studies*** in August 2024, advancing towards structured RWE use in pricing and reimbursement decisions.[2]
	HAS requires ***thorough documentation of methodologies*** to ensure reliable RWD for the French healthcare contexts, as detailed in its 2021 guidance.[3]
	Alongside MHRA's regulatory integration of RWD, NICE has shown strong commitment to RWE by publishing the ***NICE RWE framework,*** embedding methodological standards into assessments and increasingly incorporating RWD into appraisals.[7]

Data platforms are essential for aggregating healthcare information from various sources for secondary purposes like research and policy-making, while ensuring compliance with privacy regulations such as general data protection regulation (GDPR) across Europe

Table 4: **Overview of health data platforms**

	Employs a decentralized, federated approach through the ***Medical Informatics Initiative***, promoting interoperability and secure data exchange while adhering to stringent data protection laws.[8]
	Utilizes the ***Valtermed system*** to collect RWD and inform decisions about high-impact medicines.[9]
	Does not have a similar centralized system for health data aggregation
	Leads with its centralized ***Health Data Hub***, enabling secure access and innovation-driven data sharing within a strict privacy framework.[10]
	The ***NHS Digital platform*** serves as a central repository, supporting service improvement and research efforts.[11]

Complementing these national initiatives, the European Health Data Space (EHDS) seeks to link existing health data platforms across Europe, enabling seamless and secure exchange of health information for research and innovation.[12]

Integrating RWD into decision-making presents unique challenges across countries due to differences in healthcare systems, infrastructure, and population characteristics

Each country must comply with the GDPR as well as its own national data protection regulations, adding complexity to RWD integration.

Table 5: **Country- specific barriers to RWD integration: national data protection regulations**

	Germany faces challenges from a ***decentralized healthcare system*** and ***strict data protection rules*** under the Federal Data Protection Act.[13]
	Italy and Spain face challenges in creating uniform RWE datasets due to ***regional data collection differences.*** Additionally, Spain's regulations overseen by the Spanish Data Protection Agency and Italy's national regulations enforced by the Data Protection Authority further complicate health data integration into research efforts.[14,15]
	France's stringent privacy laws, such as the ***French Data Protection Act***, hinder institutional data sharing and contribute to regional variability, impacting the representativeness of RWE studies.[16]
	The UK has implemented its own version of GDPR (***UK GDPR***) and the ***Data Protection Act 2018*** post-Brexit, further complicating compliance for RWE studies.[17,18]

Conclusions

Despite advances, legal, infrastructural, and methodological barriers persist. Clinical trial evidence still dominates HTA decisions. Greater harmonization, via initiatives like EHDS and the new launched GREG, is needed to standardize RWD use and enable cross-border collaboration for improved patient-centered healthcare decision-making in Europe.

References

1. Kent S, et al. Clin Pharmacol Ther. 2025;117:961–6
2. Robert Giovanni Nisticò. AIFA. 2024
3. DEMESP cross-disciplinary project team. HAS. 2021
4. Bundesgesetzblatt Jahrgang 2024 Teil I Nr. 102, ausgegeben zu Bonn am 26. März 2024
5. Health Data Lab. 2025
6. AEMPS. Strategic Plan 2023-2026

See QR code for full list of references

Poster presented at GetReal® 2025 Conference, 1-2 October – Real-World Evidence with Real-World Confidence: Shaping the Future of Healthcare
Utrecht, The Netherlands

briya welcomes **AIRE**, the first AI Research Environment for Medical Innovation

Optimizing EMR Data Completeness with On-Premise AI: A Study on LLM Enhancing Medication Documentation

Dr Shaked, MD MPH, Talia Kustin, PhD, David Gruzman, Siegal Sadetzki, MD MPH | Briya, Tel Aviv, Israel

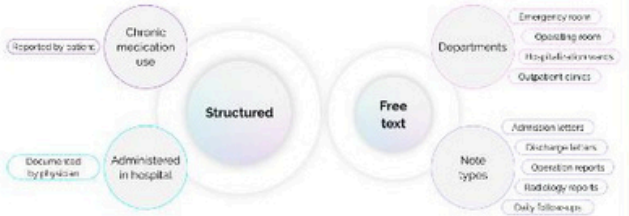
Try out AIRE by briya

Background and Objectives

Medication exposures such as aspirin during pregnancy may influence the risk of postpartum hemorrhage and other obstetric complications. However, documentation of aspirin intake is frequently captured only in unstructured free-text fields (e.g., anamnesis) within electronic medical records (EMRs), limiting data accessibility for research, contributing to incomplete medication records, and perpetuating knowledge gaps in obstetric pharmacovigilance.

This study aims to evaluate the completeness of structured EMR medication data and assess the potential of large language models (LLMs) for extracting aspirin intake information from free-text clinical notes to enhance research validity and data completeness.

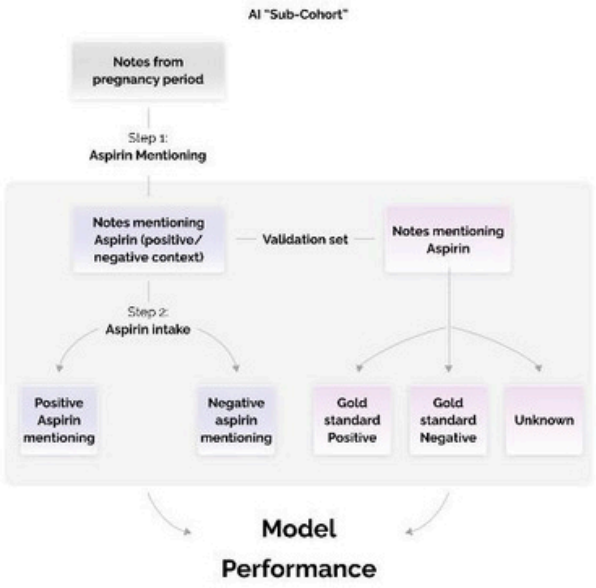
How medication use data is stored in EMRs



Methods

Data on medication use (Acetylsalicylic Acid, "Aspirin") was derived from available EMR sources, coded and free-text fields in several hospitals within Briya's network. Medication listed in coded fields was identified using ATC codes. Free-text analysis of EMR from the documenting departments (emergency room, delivery room, high-risk and post-delivery) was performed using on-premise open-source LLM (LLama 3-8B). Frequency of aspirin use was described by source of data, department and hospital. Sample of notes were reviewed by human experts for validation of LLM and F1 score, precision and sensitivity were calculated.

Briya NLP Model Pipeline



Negative Mentioning Analysis

Type 1

Recommended but did not take

Examples:
- "Aspirin was recommended, did not start yet"
- "Not taking Aspirin, refused in her first pregnancy"

Type 2

Is not recommended

Examples:
- At current status, Aspirin use is not recommended
- To my estimate, Aspirin is not recommended
- No Aspirin use during pregnancy

LLM on free-text revealed patient treatment aspects that cannot be inferred from structured data

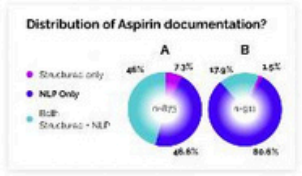
Results

The study population included 16,122 women who gave birth to 16,185 infants (18,836 deliveries) between 03.2020 - 05.2024. 1,142 women (6.1% of deliveries) women used aspirin during pregnancy. Positive aspirin use was documented 1,596 times, 1,181 in free-text (range 1-6 notes/women). Compared to experts, the validation of LLM were: F1=0.9, precision=0.88, recall=0.92. About 66% of patients who received the medication (n=755) were identified exclusively via free-text, another third (n=358) were identified by both free-text and coded fields, resulting in free-text coverage of 97.3%. The remaining 2.7% of aspirin uptake was identified exclusively in coded fields. Comparison by documenting department and hospitals, dosage, and treatment duration will be presented.

How medication use data is stored in EMRs

	Hospital A (n= 18,841 deliveries)		Hospital B (n= 26,849 deliveries)	
	n	%		% of deliveries
Singleton	18,495	98.17	26,539	98.83
Twins	341	1.82	307	1.15
Triplets	3	0.01	3	0.01
Mother age, years	n= 16,101 mothers		n= 22,952 mothers	
Mean (std)	30.4 (5.2)		29.5 (5.4)	
Range	15.2 - 55.0		16.5 - 54.4	
Delivery week				
Mean (std)	38.1 (2.2)		39.4 (1.5)	
Range	23.4 - 43.4		23.2 - 43.8	

	Sensitivity (Recall)	Precision
Hospital A	0.93	0.98
Hospital B	0.96	0.91



Conclusions

AI-powered free-text abstraction has the potential to capture the majority of clinically relevant medication data that might otherwise be missed in structured records. Large language models (LLMs) exhibit outstanding zero-shot performance, accurately extracting medication information from unstructured clinical notes without the need for task-specific training. Notably, these models can be deployed on-premise, without reliance on GPUs, thereby efficiently optimizing hospital IT resources. The completeness and quality of medication-use data remain highly dependent on both the original documentation and its source within the EMR.

briya welcomes **AIRE** the first AI Research Environment for Medical Innovation

Unlocking the Potential of Radiology Reports Using NLP: A Real-World Data Approach to Rotator Cuff Tear Severity

Dr Shaked, MD MPH | Siegal Sadetzki, MD MPH | Talia Kustin, PhD | Ben Giladi, MD | Chen Patt, MD | Talia Tron, PhD | Briya, Tel Aviv, Israel

Objectives

Rotator cuff (RC) tears present significant clinical challenges, with treatment decisions—ranging from conservative management to surgical intervention—relying heavily on accurate **tear severity assessment**. However, International Classification of Diseases (ICD) codes often lack granularity to capture RC tear severity, complicating the use of **real-world data (RWD)** for research and clinical decision-making. This study aimed to utilize **machine learning (ML) Natural Language Processing (NLP)** to extract and classify RC tear severity from radiology reports, enabling scalable and precise identification of massive RC tears in RWD sources.

Methods

To address this, we developed an **on-premise** predictive model that categorizes **massive rotator cuff tears (MRCTs)** into three distinct types (**Table T1**) using hospital-collected shoulder MRI reports. Given computational constraints in clinical environments, we prioritize **memory efficiency** and **explainability** by employing a **hybrid** approach:

- **Rule-based extraction (regex)**
  Identifies key terms like "full-thickness", "supraspinatus", or "retraction" from reports
- **Classical NLP models (trainable on limited hardware)**
  Analyze structured features such as tear size, involved tendons, and atrophy descriptors
- **LLM integration (LLaMA 8B, deployed locally)**
  Generates contextual interpretations of ambiguous findings (e.g., "extensive tendinous involvement")

Table T1: **Definition of Massive Rotator Cuff Tear (MRCT):**
Any of the following conditions:

1 Specific mentioning of **massive** RC tear

2 Complete tear measuring ≥5 cm in any of the four RC tendons

3 Complete rupture of two or more of the four RC tendons

Shoulder MRI reports were collected from a single hospital system. Clinical experts annotated **623 MRI reports** for massive rotator cuff tear categories (**Table T1**). **147 reports (833 sentences)**, were further annotated at the **sentence level**, for tear type mentions (complete, massive, partial, no tear, cannot infer) and specific muscle **involvement** (supraspinatus, subscapularis, infraspinatus, teres **minor**).
These were used for the **Briya© NLP model training**.

We developed a **hybrid pipeline**. MRI reports were first scanned for **relevant keywords**, then analyzed by **specialized classifiers**. The **AI model** provided additional insights, and results were merged using predefined rules to determine if a tear was 'massive' (***Table T2** and **Figure A***). This approach balances **accuracy** with the ability to run **securely** within hospital systems while maintaining **transparent decision-making**. Performance was assessed separately for **individual NLP models** and the **integrated pipeline** using standard classification metrics (**accuracy, precision, f1score**).
Clinical experts reviewed outputs to validate clinical relevance and explainability of predictions.

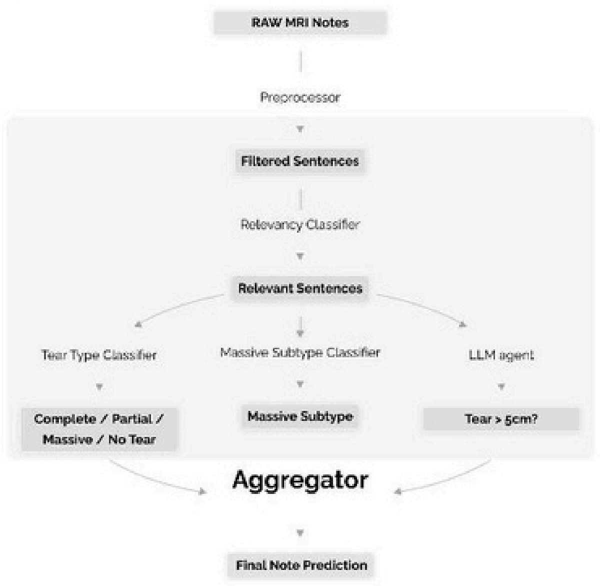
Table T2: **Briya NLP Model Pipeline**

Pipeline Stage	Description	Technical Details
Text Preprocessing	Segments MRI reports into sentences and filters content using shoulder pathology keywords (see below)	Hebrew token normalization, split to sentences keyword filtering*
1. Relevancy classifier	Uses ML to classify sentences as clinically relevant/irrelevant for rotator cuff tear descriptions	Binary classifier**
2. Tear Type Categorization	Classifies eligible sentences into discrete rotator cuff tear categories	Multiclass classifier** to **one of the** following: complete/massive/partial/no tear/indeterminate
3. Massive Tear Subtyping	Refines massive tear predictions using LLMs and NLP for contextual analysis	Multilabel classifier** to **any of the following** (complete, massive 1, massive 2)
Result Aggregation	Synthesizes predictions across all stages into unified report-level classification	Aggregation logic combining pipeline outputs

* Keywords filter for task: [illegible]
** [illegible]

Briya NLP Model Pipeline

Figure A:



LLM* to identify tears larger than 5cm:

```
massive_tear2_template = """Does the following text mention a shoulder tear larger than 5 cm?
INSTRUCTIONS:
- If a shoulder tear size is mentioned ([illegible]) - return "larger than 5cm" or "smaller than 5cm"
- If no shoulder tear is mentioned - return "not mentioned"
- If a shoulder tear is mentioned but with no explicit size - "not mentioned"

Response should be in the following format:
explanation = short explanation of the conclusion
result = "larger than 5cm" \ "smaller than 5cm" \ not mentioned

TEXT:
[illegible]
}
"""
```

Hebrew Normalization:

```
def normalize_hebrew(text):
    # 1. Remove Hebrew diacritics (Niqud and Taamim)
    text = re.sub(r"[\u0591-\u05C7]", "", text)

    # 2. Remove non-informative prefixes (adjust as needed)
    text = re.sub(r"[illegible]", "", text)

    # 3. Remove non-informative suffixes (adjust as needed)
    text = re.sub(r"[illegible]", r"[illegible]", text)

    # 4. Remove unwanted characters (keep numbers, Hebrew, English)
    text = re.sub(r"[illegible]", "", text)

    # 5. Normalize whitespace
    text = re.sub(r"\s+", " ", text).strip()

    return text
```

Results

Radiology reports of **406 patients** were analyzed,
with a **mean age of 58.8 years** (standard deviation=8.78),
and 61.0% of the patients were male.
56.4% of the scans were of the right shoulder and 43.6% of the left shoulder

Data preprocessing:

Report Level:
A total of 623 reports were annotated, with 619 reports (99%) containing relevant keywords. Based on clinical expert tagging, **97 cases (15.5%) were identified as massive rotator cuff tears.**

Sentence Level:
Reports consisted of 24,690 sentences in total, with 6,392 (25%) relevant sentences after filtering related keywords. The mean sentence per report was 20 with standard deviation of ±5 sentences.

Data Descriptions:

We used a dual annotation strategy: Combines **document-level classification** with **fine-grained sequence labeling (by sentence)**.

Training: Sentence-level annotations
- 147 reports with sentence-level annotations (833 sentences) used for model training
- Used for training sequence labeling models

Evaluation: Full-report annotations
- 623 reports fully annotated for report-level classification (e.g., 15% massive tears)
- 99% of reports contain relevant keywords
- Average report length: 20 ± 5 sentences

Sentence-level annotations
Label distribution by sentence:

Label	Count	Proportion
For Sure Complete*	321	0.38
Not Related	263	0.31
Partial	143	0.17
No Tear	47	0.05
For Sure Massive_1	25	0.03
For Sure Massive_2	9	0.01
Other	37	0.04

* [illegible]

Briya NLP model:
Results

Pipeline Stage	Description
Text Preprocessing	• Notes from 623 MRI reports - 527 not massive, 97 massive • Split to train (60%) and test (40%) for training and evaluation • Split to 24,690 sentences • 25% (6,392) sentences filtered by relevant keywords
Relevancy classifier	• Average Precision: 91% • Average recall: 96%
Final note prediction	• Average Precision: 91% • Average Recall: 90%

Conclusions

This RWD-driven NLP approach offers a robust, scalable solution for identifying and classifying massive RC tears from MRI radiology reports, addressing the limitations of severity assessment of ICD coded conditions in real-world healthcare settings.

By enabling enhanced data extraction and improving the precision of RC tear diagnosis, this method has the potential to support both clinical decision-making and large-scale RWD research on RC tears.

briya welcomes **AIRE**, the first AI Research Environment for Medical Innovation

Leveraging RWD and NLP to Identify At Risk Metabolic Dysfunction Associated Steatohepatitis (MASH)

Dr Shaked, MD MPH, Talia Tron, PhD , Gadi Lalazar, MD | Briya, Tel Aviv, Israel

Try out AIRE by briya

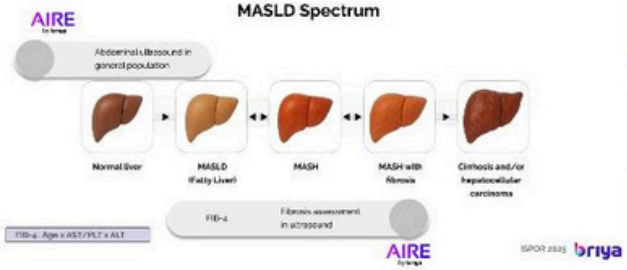
Background and objectives

Affecting approximately 32% of the adult population worldwide (https://e-cmh.org/journal/view.php?doi=10.3350/cmh.2022.0369). MASH represents the more aggressive form within the spectrum of metabolic dysfunction-associated steatotic liver disease (MASLD)

In spite of its prevalence and significant disease burden, MASLD and its progressive forms remain largely under-documented and undiagnosed in patients' records. The clinical importance is to identify the condition in early stages in order to slow down disease progression.

This research aims to identify undiagnosed or undocumented cases of fatty liver disease and improve early detection of MASH by stratifying at-risk patient populations. Segmenting these populations enables more targeted screening, improves tracking of disease progression, and increases the likelihood of early detection, timely follow-up, and appropriate treatment.

The goal is to leverage existing data from EHRs, laboratory tests, and radiology reports by training NLP models to support the identification of relevant patient populations, thereby reducing overhead and streamlining research workflows.



Methods

We analyzed abdominal radiology reports from patients with at least one abnormal liver function test (LFT) between 2020-2024, collected from a leading state-mandated health provider. Reports were labeled using multiple validation sources, including expert radiologist review, coded diagnoses, and LFTs. We developed a machine learning NLP classification model using the Briya® computational platform (Briya NLP). Model performance was assessed using area under the curve (AUC), sensitivity, specificity, and accuracy metrics.

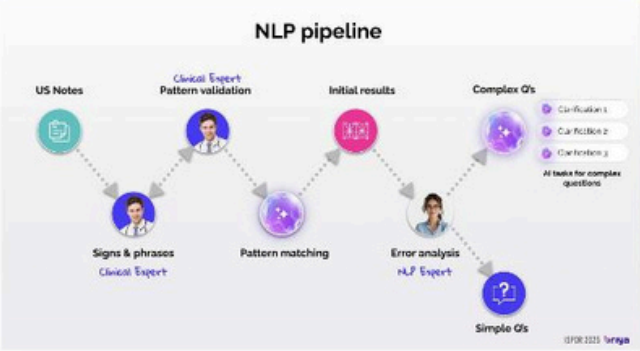
This retrospective study was conducted using EMR, lab, and radiology data accessed, de-identified, and standardized from across several partnering hospitals using Briya AIRE. The patient cohort was defined by analyzing abdominal radiology reports from patients with at least one abnormal liver function test (LFT) between 2020-2024. We developed an NLP model to identify signs of fatty liver in ultrasound reports, using simple pattern-matching techniques based on common terms and phrases typically found in clinical documentation ("fatty liver" or "fat infiltration").

As a next step we stratified the population into a high-risk and low-risk group. We used existing lab result data found in EHRs, age, and gender to calculate FIB-4 scores (FIB-4= Age x AST/PLT x ALT). For patients with multiple FIB-4 scores, we selected the value closest in time to the date of fatty liver diagnosis.

Since fibrosis risk is not explicitly mentioned in clinical notes, a simple NLP pattern-matching approach proved insufficient for identifying at-risk patients. To address more complex indicators of fibrosis, we trained the model on abdominal ultrasound reports to detect and interpret mentions of spleen size, liver morphology and dimensions, varices, ascites, and other relevant features.

This enabled the generation of an additional MASH risk score ("Briya FIB Score") which was then combined with the FIB-4 results. Based on this integrated approach, patients were stratified into high-risk and low-risk groups.

Briya NLP Model Pipeline



Simple pattern matching

Precision	Recall
0.6	0.6

Briya AIRE

Precision	Recall
0.77	0.94

Results

This analysis covered 28,795 radiology reports taken between 2020 - 2024 from 20,422 unique patients with abdominal ultrasound and identified an additional 4,036 patients with confirmed fatty liver diagnosis. The size of the relevant study population increased by 90% (from 1,122 to 5,158). The analysis revealed that only 22% (1,122 of a total of 5,158) of MASLD patients had a clearly documented diagnosis in structured EHR fields. In contrast, 85% (4,395 of a total of 5,158) of the total patient cohort had a fatty liver diagnosis visible in ultrasound reports, with 78% identified exclusively through unstructured ultrasound data. 7% (359 of 5,158) could be identified using both ICD-9 code and US-analysis.

Conclusions

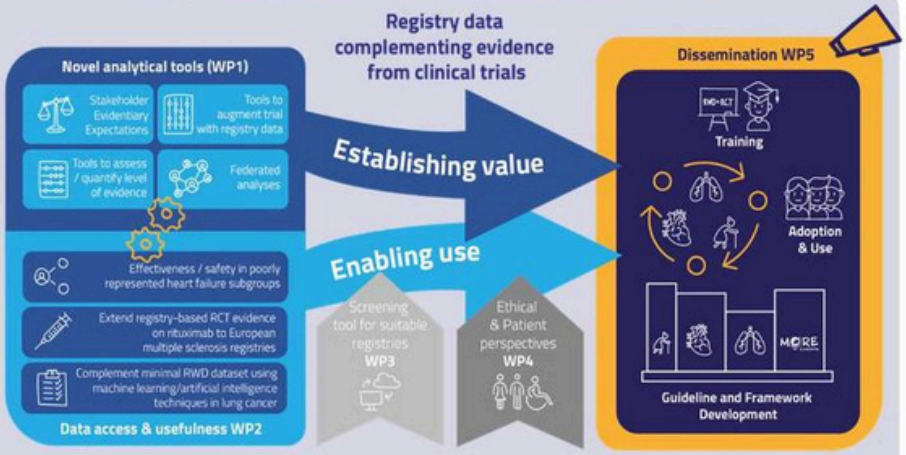
Our RWD-driven NLP based approach provides an efficient, scalable solution to identify under documented cases of fatty liver in routine clinical practice. This automated system could enhance earlier intervention through lifestyle modifications and targeted therapies, ultimately improving patient outcomes in real-world healthcare settings.

MORE EUROPA

More Effective and ethical Use of Registry data to suppOrt PAtient-centred regulatory and HTA decision making

There is a lack of knowledge on how registry-based real-world data (RWD) can complement randomized controlled trial (RCT) data to provide more robust estimates of the efficacy and safety of medicinal products across all subgroups of European patients.

Six work packages work towards a more efficient use of RWD for the development, registration and assessment of medicinal products in Europe



We aim to develop, implement and establish evidentiary standards and methods to address the data and evidentiary needs of regulatory authorities and health technology assessment (HTA) bodies

Case Studies

Heart Failure	To enable use of RWD to assess effectiveness and safety in subgroups of patients who are poorly represented in randomized controlled trials (RCTs)
Multiple Sclerosis	To extend registry-based RCT evidence supporting the use of rituximab by using RWD from different European multiple sclerosis registries for effectiveness and cost-benefit
Lung Cancer	To complement the minimal registry dataset using machine learning/ artificial intelligence

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.

Université Paris Cité · quinten health · Karolinska Institutet · CHARITÉ · Radboudumc · DICA · umcg · UPPSALA UNIVERSITET · CRITICAL PATH INSTITUTE · emsp EUROPEAN MULTIPLE SCLEROSIS PLATFORM · EURORDIS RARE DISEASES EUROPE · LISBOA · FACULDADE DE FARMÁCIA · IQVIA · UNIVERSITÀ CATTOLICA del Sacro Cuore · quinten

Impact

- Decrease the costs associated to drug development/ licensing
- Empowering patients through ethical and transparent decision-making
- Speed up the accessibility and reimbursement of drugs to European people in need
- Transforming health systems for long-term sustainability and innovation

Visit our website: watch the webinars & subscribe to the newsletter here!

Contact us

more-europa@umcg.nl

LinkedIn

MORE EUROPA

Using Large Language Models for Scalable Extraction of Real-World Progression Events Across Multiple Cancer Types

Aaron B. Cohen, MD, MSCE[1,2]; Konstantin Krismer, PhD[1]; Kelly Magee, MS, FNP-BC[1]; James Gippetti, MS[1]; Aaron Dolor, PhD[1]; Tori Williams, BA[1]; Erin Fidyk, MSN, MBA[1]; Hankyul Kim, MS[1]; Qianyu Yuan, PhD[1]; Melissa Estevez, MS[1]

[1]Flatiron Health, New York, NY; [2]Department of Medicine, Grossman School of Medicine, New York University, NY

E-poster and Supplement

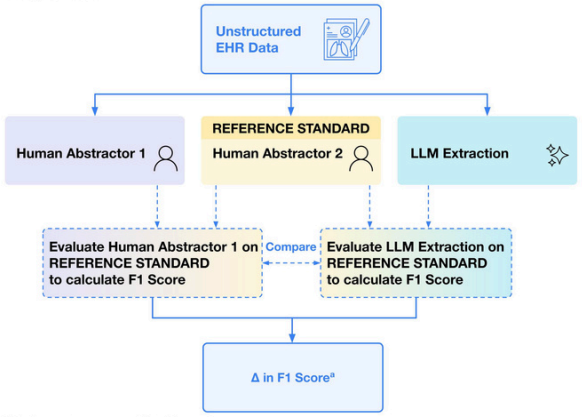
Background

- Accurate identification of cancer progression events from electronic health records (EHRs) can help generate high-quality real-world evidence (RWE) and enable promising oncology applications such as predicting disease trajectory and assessing treatment efficacy
- These use cases require both large-scale and high-quality data, but manual abstraction of real-world progression (rwP)[1,2] is time-intensive, difficult to scale, and inherently challenging given the varied and unstructured ways it can be documented across cancer types
- Large language models (LLMs) offer a scalable alternative, but their ability to extract complicated endpoints as well as their accuracy relative to expert-human abstractors is unclear
- We comprehensively evaluated the ability of LLMs to extract rwP events and their associated dates across 14 different types of advanced cancer
- To better contextualize the LLM's ability to extract rwP relative to an expert-human abstractor (and potentially identify situations where an LLM may even outperform a human), we assessed performance of the LLM curation approach against a duplicate human-abstracted reference standard (**Figure 1**)[3]
- We also assessed how using LLM-extracted data impacted real-world progression-free survival (rwPFS) estimates compared to using expert-human abstracted data

Methods

1. We applied LLM-based extraction techniques to unstructured EHR text for 14 types of advanced cancer from the Flatiron Health Research Database[4] (**Table 1**)
2. Various prompt engineering strategies including zero-shot, few-shot, and chain-of-thought were tested to optimize LLM extraction performance
3. For each cancer type, we assessed the following metrics (**Table 1**, **Figure 2**):
   a. Completeness of rwP events [the percent of patients for which the LLM extracts at least 1 rwP event following first-line (1L) therapy start date]
   b. Agreement between the LLM and expert-human abstractor on the first rwP date (± 30 days) in the 1L setting
   c. Comparison of rwPFS calculated from LLM-curated data versus expert-human abstractor-curated data, indexed to 1L therapy start date
4. We also assessed the difference in performance (Δ in F1 score) between LLM and expert-human curation approaches when compared to a second expert human abstracted reference standard (**Figure 1**, **Table 1**)

Figure 1. Evaluation Process for Assessing LLM-Extracted Real-World Progression



[a]Performance across all rwP events

Results

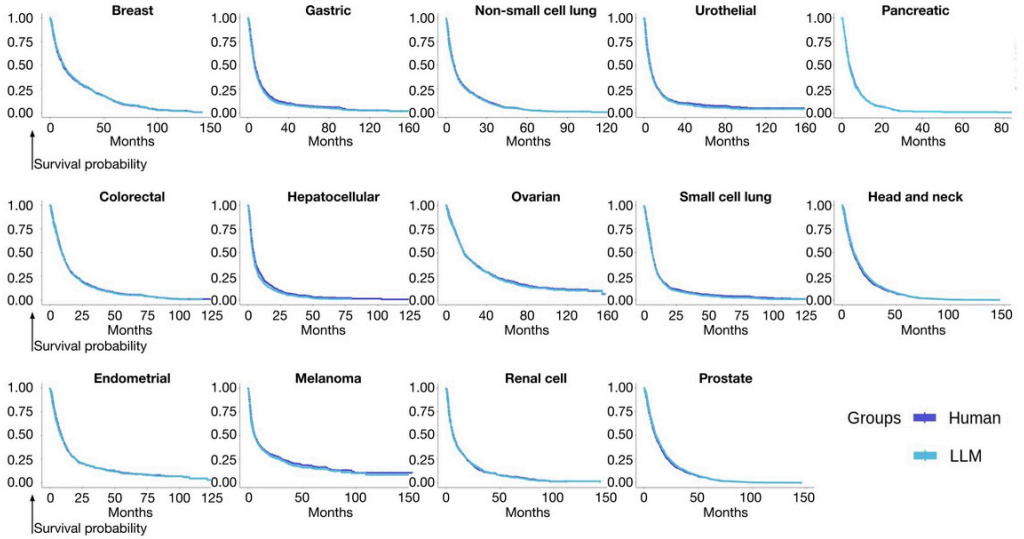
- Across all cancer types, agreement between the LLM and expert-human abstractor on the presence of at least 1 rwP event ranged 85%-91% while first rwP date agreement (± 30 days) ranged 77%-90% (**Table 1**)
- The difference in F1 score between the LLM and expert-human abstraction for majority of cancer types (10/14) was 0.7-7.8 (**Table 1**)
- The difference in F1 score for endometrial cancer, melanoma, and hepatocellular carcinoma was 9-11 points, driven by lower precision as a result of model difficulty differentiating between advanced diagnosis workup and early progression events (**Table 1**)
- A comparison of rwPFS between the LLM and expert-human abstractors showed <1 month difference in median rwPFS estimates and overlapping 95% CIs across all 14 cancer types (**Figure 2**)

Table 1. Performance Metrics for LLM-Extracted rwP

	Completeness[a]	Agreement between LLM and Human[a]		Curation evaluation on Reference Standard
Cancer type	Patients with rwP event (%)	Presence of rwP (%)	First rwP date (%)	Δ in F1 score[b]
Breast	72.7	88.4	80.6	5.6
Colorectal	75.7	88.9	81.8	5.5
Endometrial	79.9	90.7	85.9	9.3
Gastric	68.6	89.7	86.2	7.8
Head and neck	92.9[c]	90.8	85.4	3.0
Hepatocellular	66.3	85.0	84.3	11.2
Melanoma	73.2	87.7	85.8	10.4
Non-small cell lung	70.2	89.7	86.8	5.1
Ovarian	63.8	90.6	86.4	6.2
Pancreatic	63.8	90.9	84.8	2.7
Prostate	98.4[c]	85.4	76.7	3.5
Renal cell	75.9	89.3	80.4	6.9
Urothelial	73.0	88.4	84.6	3.6
Small cell lung	70.1	90.8	90.1	0.7

[a]Completeness and Agreement metrics based on 1000 patients per disease with exception of head and neck (N = 850), pancreatic (N = 497), and prostate (N = 1513). [b]Counts to calculate F1 metrics available in supplementary table via QR code. [c]Cohorts enriched for patients with >1 line-of-therapy, increasing prevalence of progression events

Figure 2. Comparison of rwPFS by Curation Approach Across Cancer Types



Median rwPFS (95% CI), mo	Breast (n = 1000)	Colorectal (n = 1000)	Endometrial (n = 1000)	Gastric (n = 1000)	Head and neck (n = 850)	Hepatocellular (n = 1000)	Melanoma (n = 1000)	NSCLC[a] (n = 1000)	Ovarian (n = 1000)	Pancreatic (n = 497)	Prostate (n = 1513)	Renal cell (n = 1000)	SCLC[b] (n = 1000)	Urothelial (n = 1000)
Expert-human abstracted	12.2 [11.0-13.8]	9.3 [8.6-10.5]	9.0 [8.0-9.8]	6.8 [6.4-7.6]	5.7 [5.3-6.1]	4.4 [3.9-4.8]	6.2 [5.5-7.8]	6.0 [5.3-6.6]	17.6 [16.2-19.7]	3.9 [3.4-4.5]	10.5 [9.6-11.3]	6.8 [6.1-7.7]	6.2 [5.8-6.9]	6.1 [5.5-6.7]
LLM-extracted	12.6 [11.5-14.7]	9.3 [8.5-10.1]	8.5 [7.6-9.6]	6.5 [6.0-6.9]	5.9 [5.4-6.3]	3.6 [3.4-4.2]	6.1 [5.3-7.2]	5.8 [5.2-6.4]	17.8 [6.3-20.7]	3.5 [3.2-4.0]	11.0 [10.3-12.2]	6.7 [6.0-7.4]	6.1 [5.7-6.6]	5.8 [5.3-6.4]

Abbreviations: mo, months; NSCLC, non-small cell lung cancer; SCLC, small cell lung cancer.

A **novel LLM** approach **accurately extracted rwP events and dates** across **14 cancers**, demonstrating its **generalizability and validity**

Conclusions and Future Directions

- LLMs were able to extract rwP, a complex and clinically important endpoint, with high performance that generalized across a multitude of cancer types
- Direct comparison of LLM extraction to expert-human abstraction was made possible by generating a second human-abstracted reference standard for evaluation across each of the 14 different cancers
- LLM performance was additionally validated by assessing interpretable metrics such as completeness, agreement, and the ability to replicate rwPFS analyses as compared to expert-human abstraction
- This comprehensive and transparent approach to EHR data curation and evaluation provides important information on the quality of a critical and nuanced endpoint to support a variety of use cases while also building trust in the data
- As LLM capabilities continue to improve, future work should explore extracting rwP across additional cancers (e.g., hematologic malignancies) as well as extracting additional clinical details (e.g., growth of existing lesions vs development of new sites of disease)
- Continued evaluation should explore drivers of misclassification and error modes, facilitating iteration of prompt engineering techniques and improvements in model performance
- Future work should explore how incorporating LLM-extracted rwP into analyses and predictive algorithms may help improve clinical decision-making, patient care, and accelerate oncology research

References

1. Torres AZ, et al. Analysis of a real-world progression variable and related endpoints for patients with five different cancer types. *Adv Ther*. 2022;39:2831-2849. doi:10.1007/s12325-022-02091-8
2. Griffith SD, et al. Characterizing the feasibility and performance of real-world tumor progression end points and their association with overall survival in a large advanced non-small cell lung cancer data set. *JCO Clin Cancer Inform*. 2019;3(3):CCI.19.00013. doi:10.1200/cci.19.00013
3. Estevez M, et al. Ensuring reliability of curated EHR-derived data: the Validation of Accuracy for LLM/ML-Extracted Information and Data (VALID) Framework. *arXiv*. Published online 2025. doi:10.48550/arXiv.2506.08231
4. Flatiron Health. Database Characterization Guide. Flatiron.com. Published March 18, 2025. Accessed June 18, 2025. https://flatiron.com/database-characterization

Acknowledgments: Darren Johnson for editorial support; Madeline Morenberg and Natalie Marcus for design

Disclosures: This study was sponsored by Flatiron Health, Inc.—an independent member of the Roche Group. During the study period, ABC, KK, KM, JG, AD, TW, EF, HK, QY, and ME reported employment with Flatiron Health, Inc. and stock ownership in Roche. Data first presented at AACR AI and ML 2025 in Montreal, QC, Canada on July 11, 2025

Author contact information: Aaron B. Cohen, acohen@flatiron.com

Ethical practices around patient registry utilization:
A survey of European stakeholders

Fabian Windfuhr[1] , Olmo van den Akker[2], Diogo Almeida[3], Bruno Sepodes[3], Carla Torre[3], Peter G.M. Mol[1,4], Sieta T. de Vries[1,4]

[1] Department of Clinical Pharmacy and Pharmacology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands
[2] QUEST Center for Responsible Research, Berlin Institute of Health at Charité – Universitätsmedizin Berlin, Berlin, Germany
[3] Faculdade de Farmácia, Universidade de Lisboa, Portugal
[4] Dutch Medicines Evaluation Board, Utrecht, the Netherlands

Introduction

With the increasing use of patient registries in decision-making on medicines, it is important to assess ethical aspects of registry use. We aimed to assess overall stakeholder views on ethical aspects of patient registry usage, and whether these views differ across stakeholder groups.

Methods

- **Target population:** European regulators, health technology assessors (HTA), payers, industry, patients, academia, healthcare professionals (HCPs)
- **Web-based survey study:** dissemination via email (1565 personalized invitations), through newsletters, and social media
- **Survey content** (22 questions): Demographics, Perspectives towards ethics aspects of patient registry usage
- **Statistical analyses:** Descriptives (absolute & relative frequencies, means, 95% confidence intervals (95% CI))

Results

We obtained 370 survey responses (**Table 1**). Information on a registry's use-and-access policy was considered most important (mean 4.2; 95% CI: 4.1-4.3) and information on funding sources least important (mean 3.9; 95% CI: 3.8-4.0; **Figure 1**). Participants on average considered 51% (scale:0-100%; 95% CI: 48-55) funding of registries by industry acceptable (**Figure 2**). Most participants preferred the use of an opt-out consent for data collection and processing by registries (51%), and for data processing by external parties (61%; **Figure 3**). 65% preferred if registries pseudonymize data for processing (**Figure 4**). For use by external parties, the preferred option was transferring anonymized data (34%; **Figure 5**).

Table 1: *Respondents' demographics*

		N = 370 (%)	
Stakeholder	Regulators	129	(35)
	HTA/payers	30	(8)
	Industry	44	(12)
	Patients	39	(11)
	Academia/HCP/Others	128	(35)
Age[1]	(years, mean/SD)	48	(±11)
Work experience[2]	(years, mean/SD)	14	(±9)
Gender[3]	Women	206	(56)
Region[4]	Northern Europe	79	(21)
	Eastern Europe	20	(5)
	Southern Europe	64	(17)
	Western Europe	183	(50)
	Multinational org.	23	(6)

[1]3 missing, [2]1 missing [3]5 missing, [4]1 missing

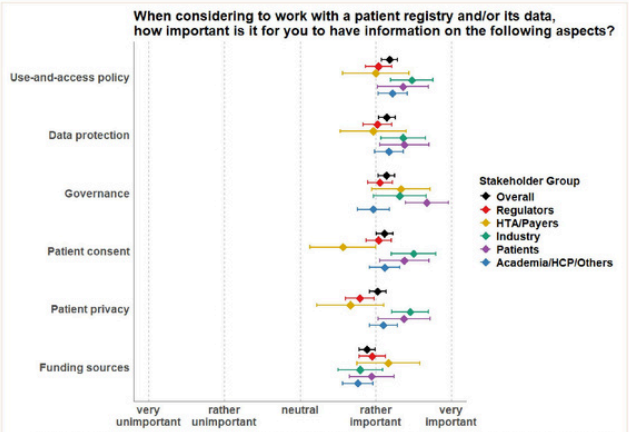


Figure 1*: Perceived importance of having information on different types of ethics-related information concerning patient registries, stratified by stakeholder group*

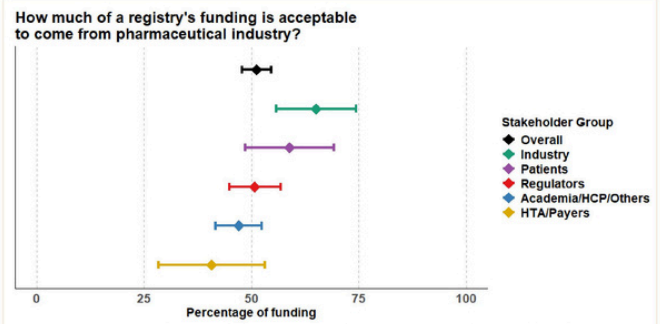


Figure 2*: Percentage of industry funding considered acceptable, stratified by stakeholder group*

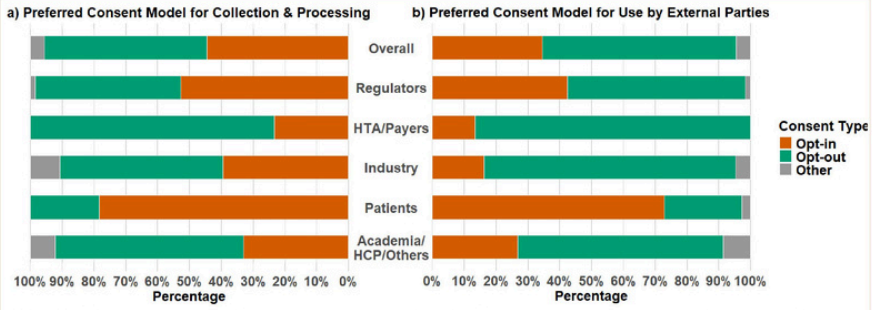


Figure 3*: Preferences for obtaining informed consent from patients for a) initial collection and processing of patient data in a registry, and b) its use by external parties, stratified by stakeholder group; relative frequencies*

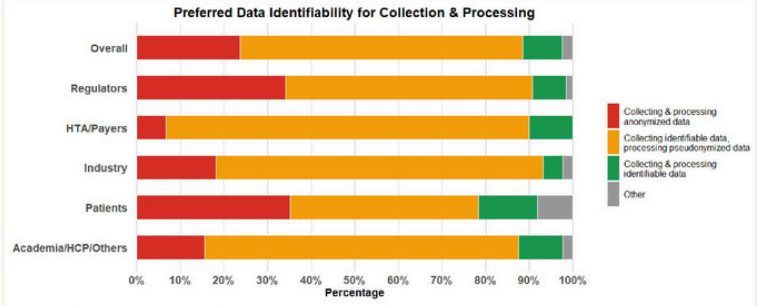


Figure 4*: Preferences on identifiability of data from enrolled patients that is collected and processed by a registry, stratified by stakeholder group; relative frequencies*

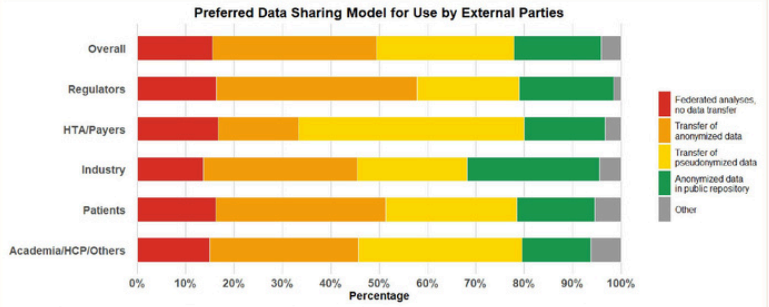


Figure 5*: Preferences on different models of data sharing to enable the secondary use of registry data by external parties, stratified by stakeholder group; relative frequencies*

Conclusions

- Stakeholders find it important to be informed on all ethical practices of patient registries assessed in this survey
- Most stakeholders prefer opt-out consent, pseudonymization for data processing, and anonymization for data sharing with external parties
- Stakeholders differed on certain views, e.g., patients prefer opt-in consent, HTA/payers prefer lower industry funding of patient registries

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.

Participate in our next survey!

f.a.windfuhr@umcg.nl | more-europa@umcg.nl

Website LinkedIn

Utrecht University | DICA DMA Dutch Medication Audit | IPOPORTO | TUMOR ZENTRUM Oberösterreich | Pharmacoepidemiology & Clinical Pharmacology

Repurposing Registries:
WHAT IS THE LEVEL OF COMPLETENESS OF CANCER REGISTRIES TO SUPPORT REGULATORY AND HTA DECISIONS?

Manon Wilpshaar[1,2]
Ansgar Weltermann[3]
Luisa Prada[4]
Doranne Hilarius[1,5]
Christine Leopold[6]

INTRODUCTION & GOAL

Cancer registries, initially designed for clinical purposes, are increasingly recognized as a data source for regulatory and health technology assessment (HTA) decision-making[1,2]. Despite developments in supporting the identification and selection of cancer registries, a translational gap remains between the registries' original purpose and their current and future expected use in policymaking. This study assesses the completeness of cancer registries regarding essential variables for regulatory and HTA decision-making and examines the evolutionary and contextual factors influencing their current suitability for such purposes.

METHODS

A case-study approach was employed by selecting three European cancer registries differing in registry type, coverage, and health system context: the Cancer Centre Upper Austria (CCUA) registry, the Dutch Medication Audit (DMA) and the Porto Comprehensive Cancer Centre (PCCC) registry. A data extraction sheet to systematically describe the registries' initiation, structure, coverage and evolution was developed and filled out in agreement with all co-authors/registry representatives. Using literature and expert input, a blueprint to assess a cancer registry's completeness in terms of regulatory and HTA needs was developed[3,4]. The blueprint will be applied to compare the completeness of the three registries.

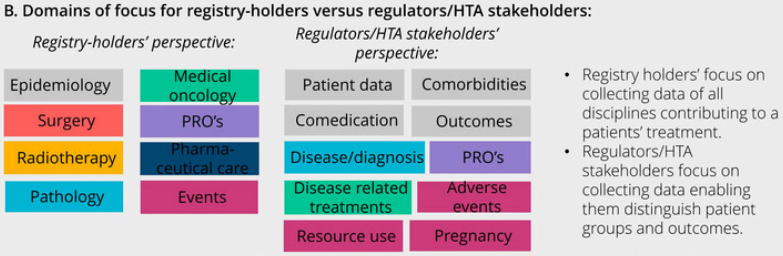
RESULTS

Although these registries were designed for clinical purposes, they exhibit clear differences in inclusion criteria driven by their specific aims. While the CCUA and PCCC registries apply a diagnosis-based inclusion, capturing all newly diagnosed cancer patients and collecting comprehensive epidemiological data, the nationwide DMA applies a treatment-based inclusion, focusing on patients receiving high-cost systemic anti-cancer therapy and between-hospital variance. The evolution of these registries is shaped by their initial aims: shifting from surgical focus to mapping the complete patient pathway.

A. Most striking differences in registry initiation, organization and coverage:

Domain	Characteristic	Cancer Centre Upper Austria	Dutch Medication Audit	Porto Comprehensive Cancer Centre
Initiation	Initial purposes	Supporting multidisciplinary tumor boards	Enabling benchmarking, identifying practice variations between hospital	Enabling epidemiological research, clinical decision-making
Organization	Owner	Network of hospitals	Dedicated institute	Single hospital
	Funding	Hospital operators	Healthcare insurers	Hospital operators
Coverage	Type of inclusion	Diagnosis-based	Treatment-based	Diagnosis-based
	Coverage of tumor entities	All tumor entities	Breast-, colorectal-, gynecological, head- and neck-, lung-, upper GI, prostate cancer	All tumor entities

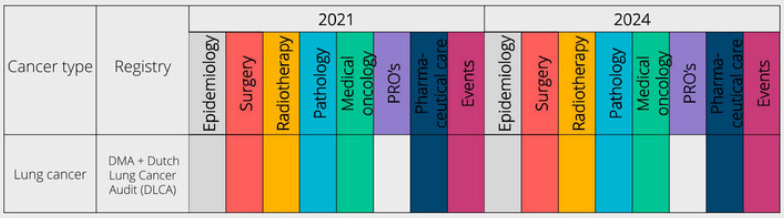
B. Domains of focus for registry-holders versus regulators/HTA stakeholders:



- Registry holders' focus on collecting data of all disciplines contributing to a patients' treatment.
- Regulators/HTA stakeholders focus on collecting data enabling them distinguish patient groups and outcomes.

Domains in registry holders' perspectives are color-coded; the domains in regulators/HTA stakeholders' perspectives are colored in the domain that has most overlap in registry holders' domains.

C. Example of visual presentation of completeness of registry over time:



Completeness will be presented from 2009-2024 in 3-year intervals; a colored cell means availability of variables in the specified domain and year; a striped colored cell means limited availability of variables in the specified domain and year.

CONCLUSIONS

The initial aims and design of cancer registries impact their ability to support regulatory and HTA decision-making. Optimally repurposing registry data requires capacity building for registry holders and regulatory/HTA stakeholders focusing on mutual understanding of aims, needs, and limitations. Additional research should address data interoperability, standardization, and data quality across registries.

NEXT STEPS

The blueprint to define completeness of cancer registries will be finalized and applied to the selected registries. Subsequently, registry completeness will be compared, and influence of initial aims and health care context will be evaluated.

GET IN TOUCH

Manon Wilpshaar, researcher at DICA and PhD-candidate at Radboud UMC

- *Email address: m.wilpshaar@dica.nl*
- *Linked-In: linkedin.com/in/manon-wilpshaar/*

REFERENCES

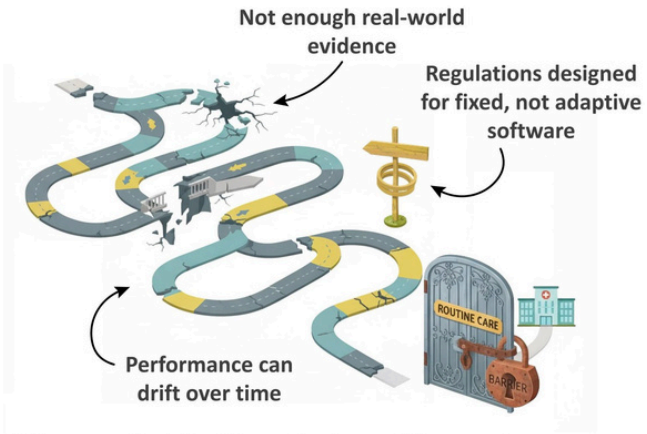
1. *Windfuhr et al. (2025)*
2. *Plueske et al. (2025)*
3. *EMA Discussion paper: Use of patient disease registries for regulatory purposes methodological and operational considerations (2018)*
4. *Zorginstituut Nederland: Eindrapportage casestudiefase Programma Regie op Registers voor Dure Geneesmiddelen (2023)*

AFFILIATIONS

1. Dutch Medication Audit, Dutch Institute for Clinical Auditing (DICA), Leiden, The Netherlands
2. Department IQ Health, Biostatistics, Radboud University Medical Center, Nijmegen, The Netherlands
3. Division of Hematology with Stem Cell Transplantation, Hemostaseology and Medical Oncology, Department of Internal Medicine I, Ordensklinikum Linz Elisabethinen, Austria
4. Porto Comprehensive Cancer Centre, Porto, Portugal
5. Department of Pharmacy, Rode Kruis Ziekenhuis, Beverwijk, The Netherlands
6. Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences (UIPS), Utrecht University, Utrecht, The Netherlands

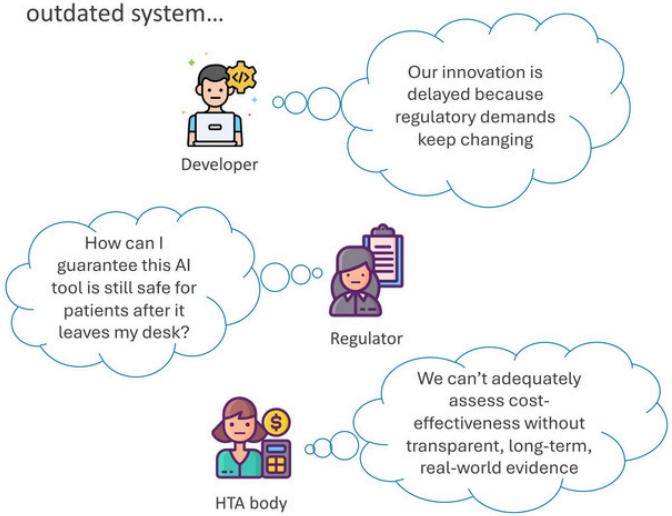
REALM: From Black Boxes to Transparent Pipelines

Why is it hard to evaluate medical AI?



This gap will delay life-saving innovations and expose health systems to biased, ineffective AI.

In addition, major stakeholders are left navigating an outdated system…



The REALM Solution

(Real-world-data Enabled Assessment for heaLth regulatory decision-Making)

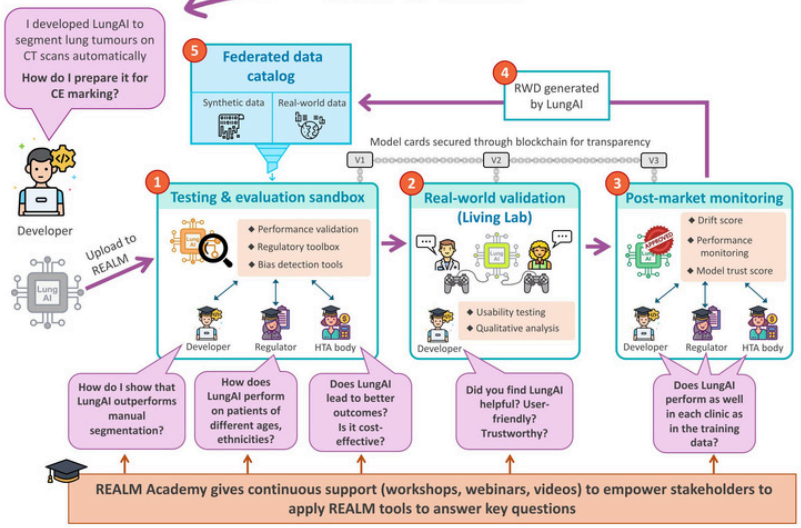
A clear and compliant path to trustworthy AI

The core framework: REALM is an **auditable**, **traceable**, and **continuously verifiable** evaluation pipeline for medical AI

The evidence backbone: We use **federated, pan-European real-world data (RWD)** to establish a robust and transparent evidence base

Regulatory alignment: REALM aligns with the forthcoming **EU AI Act** and **MDR/IVDR regulations**, giving key stakeholders one compliant route from sandbox to clinic

How it works



Impact

Safer, fairer AI across patient populations

Achieved though the Regulatory Toolbox & Post-Market Monitoring

Transparent regulatory dossiers and faster CE marking

Using standardized model cards and blockchain

Robust multi-country evidence without data pooling

Made possible by the Federated Data Catalog

EU-wide upskilling and shared evaluation culture

Driven by the REALM Academy and the Living Lab

Get involved!

Free to attend. Space is limited!

REALM Academy Workshop:
Multi-Stakeholder Insights for Evaluating Medical AI

SAVE THE DATE
5th November 2025
Location: Brussels, Belgium

Contact & Connect

a.ankolekar@maastrichtuniversity.nl

Visit us online:

realm-ai.eu | contact@realm-ai.eu | REALM_EU

Anshu Ankolekar[1], Philippe Lambin[1], Gökhan Ertaylan[2], Michel Dumontier[3]

[1] Dept. of Precision Medicine, GROW Research Institute for Oncology & Reproduction, Maastricht University, Maastricht, the Netherlands
[2] Environmental Intelligence Unit, Flemish Institute for Technological Research (VITO), Mol, Belgium
[3] Institute of Data Science, Faculty of Science and Engineering, Maastricht University, Maastricht, the Netherlands

Funded by the European Union
EU Project n°101095435

UK Research and Innovation

Characterization of Novel Longitudinal Oncology Real World Datasets in Germany and the UK

Blythe Adamson[1], Elsie Horne[2], Christoph Buhl[3], Amit Samani[2], Nicole Schinwald[3], Robin Linzmayer[3], Kathi Seidl-Rathkopf[3], Arun Sujenthiran[2]

[1]Flatiron Health, New York, NY; [2]Flatiron Health UK, London, UK; [3]Flatiron Health Germany, Berlin, Germany

Conflict of Interest Statement: This study was sponsored by Flatiron Health, Inc.—an independent member of the Roche Group. During the study period, all authors reported employment with Flatiron Health, Inc. and stock ownership in Roche.

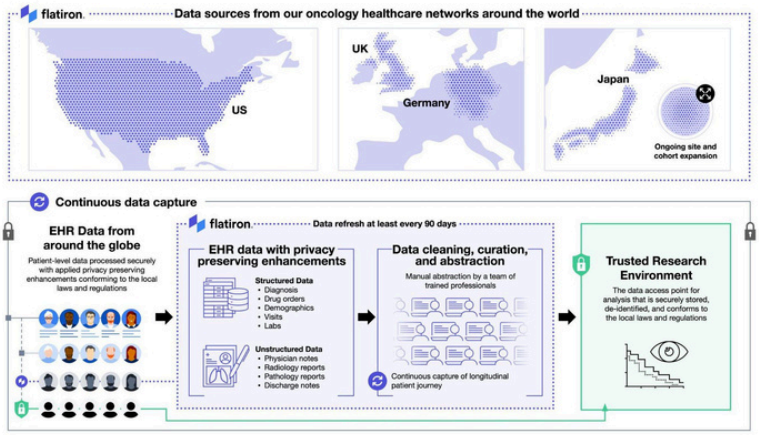
Objective

Our objective was to identify cohorts of patients with cancer in Germany and UK for longitudinal observational studies (retrospective and prospective), define curation of variables and outcomes for harmonization across countries, and to describe the prevalence rates for positive biomarker test results.

Background

- Real world oncology data with the clinical depth and recency to conduct rigorous pharmacoepidemiology studies in Europe is limited.
- Disease-specific common data models allows for harmonisation and pooling across countries in order to enhance understanding of how treatment patterns and outcomes may differ across countries where clinical guidelines and standard of care treatments differ.
- Molecular testing, such as biomarker testing and the presence of genotypic variations, is increasingly used for clinical decision making.

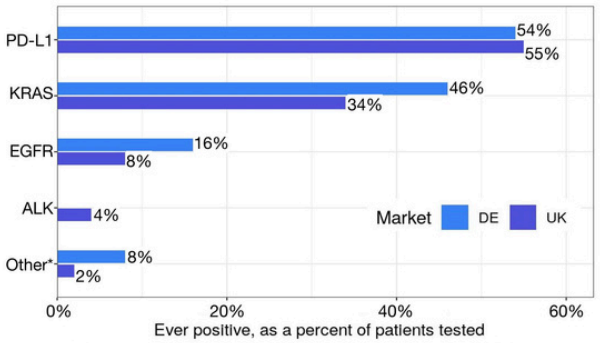
Methods



- Patients were eligible if they had a diagnosis of breast cancer or non-small cell lung cancer (NSCLC) after 1 Jan 2016 and excluded if age <18 years at time of diagnosis. The source was Flatiron Health HorizonDB, which includes 3.8 million patients electronic health record (EHR)-derived data from around the globe including the US, Germany, UK, and Japan. This study included Germany and UK patients.
- Prospective longitudinal follow-up began on 1 Jan 2024 after pre-specification of variable definitions. New relevant information is re-abstracted with 90-day recency by recurring EHR access from sites.
- The prevalence of biomarker positivity is defined as the percentage of people whose most recent test result was positive among those who were tested.

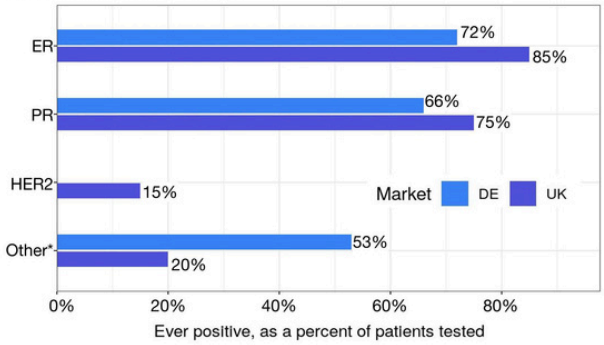
Results (continued)

Figure 2. Prevalence of biomarker positivity among people with a biomarker testing result in NSCLC. *Note that the numbers of NSCLC patients testing positive for ALK and breast patients testing positive for HER2 in the Germany dataset were <10 so have been redacted from the plots.*



*Other is any of: HER2, MET, NTRK, RET, BRAF, ROS1

Figure 3. Prevalence of biomarker positivity among people with a biomarker testing result in breast cancer.



*Other is any of: BRCA, NTRK, PD-L1, PIK3CA

Results

- The study population included 3,395 people in Germany (n = 234) and the UK (n = 3,161) diagnosed with NSCLC or breast cancer. Solid tumour testing of biomarkers found similar prevalence rates of actionable mutations and rearrangements in both countries.
- We found similar biomarker positivity prevalence rates in lung and breast cancer between UK and Germany.
- Clinically meaningful breast cancer subtypes among those with a test result were HR+/HER2- 75%, HR-/HER2- 10%, HR+/HER2+ 10%, and HR-/HER2+ 4.7%.

Table 1: NSCLC

Characteristic	Overall, N = 1,813[1]	DE, N = 169[1]	UK, N = 1,644[1]
Age at diagnosis, median [IQR]	72 (64, 77)	68 (61, 74)	72 (65, 78)
Age group			
<64	487 (27%)	65 (38%)	422 (26%)
65-74	709 (39%)	65 (38%)	644 (39%)
75+	617 (34%)	39 (23%)	578 (35%)
Sex			
F	887 (49%)	83 (49%)	804 (49%)
M	926 (51%)	86 (51%)	840 (51%)
Histology			
Squamous cell carcinoma	590 (33%)	39 (23%)	551 (34%)
Non-squamous cell carcinoma / not otherwise specified	1,223 (67%)	130 (77%)	1,093 (66%)
Smoking status			
History of smoking	1,365 (75%)	122 (72%)	1,243 (76%)
No history of smoking / smoking history unknown	448 (25%)	47 (28%)	401 (24%)
Clinical group stage			
Stage I	451 (26%)	20 (14%)	431 (28%)
Stage II	179 (10%)	26 (18%)	153 (9.8%)
Stage III	390 (23%)	41 (28%)	349 (22%)
Stage IV	688 (40%)	61 (41%)	627 (40%)
Unknown	105	21	84
Pathological group stage recorded	473 (86%[2])	77 (100%[2])	396 (84%[2])
Resection performed	549 (30%)	77 (46%)	472 (29%)
Metastatic disease at diagnosis	689 (38%)	61 (36%)	628 (38%)

[1] Median (IQR); n (%)
[2] As a % of patients with a resection performed

Table 2: Breast Cancer

Characteristic	Overall, N = 1,582[1]	DE, N = 65[1]	UK, N = 1,517[1]
Age at diagnosis, median [IQR]	64 (52, 74)	71 (54, 77)	63 (52, 74)
Age group			
<50	296 (19%)	12 (18%)	284 (19%)
50-64	542 (34%)	14 (22%)	528 (35%)
65-74	379 (24%)	17 (26%)	362 (24%)
75+	365 (23%)	22 (34%)	343 (23%)
Laterality			
Left	804 (51%)	32 (52%)	772 (51%)
Right	769 (49%)	30 (48%)	739 (49%)
Unknown	9	3	6
Menopausal status			
Pre or perimenopausal	239 (29%)	11 (39%)	228 (29%)
Postmenopausal	584 (71%)	17 (61%)	567 (71%)
Unknown	759	37	722
Tumour grade			
Grade 1-2	1,140 (73%)	37 (61%)	1,103 (74%)
Grade 3	414 (27%)	24 (39%)	390 (26%)
Unknown	28	4	24
Histology			
Invasive/Infiltrating ductal carcinoma (IDC)	1,130 (72%)	49 (80%)	1,081 (72%)
Other	434 (28%)	12 (20%)	422 (28%)
Unknown	18	4	14
Clinical group stage			
Stages I-III	203 (57%)	22 (55%)	181 (57%)
Stage IV	156 (43%)	18 (45%)	138 (43%)
Unknown	1,223	25	1,198
Pathological group stage recorded	1,151 (90%[2])	32 (73%[2])	1,119 (91%[2])
Resection performed	1,274 (81%)	44 (68%)	1,230 (81%)
Metastatic disease at diagnosis	156 (9.9%)	18 (28%)	138 (9.1%)

[1] Median (IQR); n (%)
[2] As a % of patients with a resection performed

Representative EHR-derived RWD in Europe was curated for the purpose of treatment **comparative effectiveness research**

Future directions

- Harmonised oncology EHR-derived datasets have been developed in Europe to enable cross-country comparisons of patients.
- Enhanced understanding of actionable mutations can inform global precision medicine strategies in Europe.

References

1. Ma X et al. MedRxiv. 2023. doi:10.1101/2020.03.16.20037143
2. Castellanos EH, Wittmershaus BK, Chandwani S. Raising the Bar for Real-World Data in Oncology: Approaches to Quality Across Multiple Dimensions. JCO Clin Cancer Inform. 2024;8(8):e2300046. doi:10.1200/cci.23.00046
3. Ramsey SD, Adamson BJ, Wang X, et al. Using electronic health record data to identify comparator populations for comparative effectiveness research. Journal of Medical Economics. Published online November 10, 2020. doi:10.1080/13696998.2020.1840113

Acknowledgments: Thank you to the presenter Sascha van Boemmel-Wegmann (Flatiron Health GmbH) and to Harlan Pittell, Philani Mpofu, and Eric Meadows (Flatiron Health) for scientific input. Data first presented at the ISPE 2024 Annual Meeting on August 27, 2024 in Berlin Germany

Author contact information: Sascha van Boemmel-Wegmann, sascha.vanboemmel-wegmann@flatiron.com

Integrating Real-world Evidence for Advanced Therapies in Inflammatory Bowel Disease into Reimbursement Submissions: Insights and Considerations from the UK

ABS-19

Neil R. Brett,[1] Garthiga Manickam,[1] Roy Mootoosamy,[2] Daniela Castano,[3] John S. Sampalis[1]

[1]Thermo Fisher Scientific, Montreal, Canada; [2]Thermo Fisher Scientific, London, UK; [3]Thermo Fisher Scientific, Waltham, MA, USA

Background

- Recent studies in inflammatory bowel disease (IBD) highlight notable differences between clinical trial populations and real-world study populations, including age, comorbidities, disease severity, and treatment responses.[1,2] This may lead to limited generalizability of clinical trial results in real-world populations.
- Based on the flexibility of real-world study eligibility criteria, real-world data (RWD) may be able to provide insights into the patient experiences and outcomes of a broader population, which is crucial for making informed decisions about the efficacy and cost-effectiveness of treatments.
- Thus, the UK's National Institute for Health and Care Excellence (NICE) has a strategy to leverage RWD to bridge the knowledge gaps that exist in reimbursement submissions.[3] However, it is not known if RWD was utilized in previous UK NICE health technology assessment (HTA) submissions for advanced IBD therapies.

Objective

- **The objective was to evaluate RWE use in previous NICE HTA submissions for advanced IBD therapies for adults, including biologics and Janus kinase (JAK) inhibitors, and offer recommendations for incorporating real-world evidence (RWE) into future HTA submissions.**

Methods

- A targeted review of NICE appraisals was performed for Crohn's disease (CD) and ulcerative colitis (UC) to understand RWE expectations, integration and usage.
- Available therapies for CD and UC were identified, focusing on those approved for adults. Appraisals for the approved therapies were retrieved from the NICE database. Each appraisal was independently reviewed by two individuals to extract information (year, indication, RWE utilization/context, and RWE study design.)

Results

- As of June 2025, eight biologics and three JAK inhibitors have been approved for adult CD and UC in Europe.

Figure 1. Approved therapies

	CD	UC
Biologics approved in adults	• Adalimumab • Infliximab • Risankizumab • Ustekinumab • Guselkumab • Mirikizumab • Vedolizumab	• Adalimumab • Golimumab • Infliximab • Risankizumab • Ustekinumab • Guselkumab • Mirikizumab • Vedolizumab
JAK inhibitors approved in adults	• Upadacitinib	• Filgotinib • Tofacitinib • Upadacitinib

Abbreviations: CD = Crohn's disease; IBD = inflammatory bowel diseases; JAK = Janus kinase; UC = ulcerative colitis

Table 1. Evaluation of NICE appraisals for therapies approved for CD and UC in adults from January 2015 to June 2025

Therapy (appraisal number)	Indication	Year of NICE appraisal	Utilization and context of RWE data
Biologics			
Adalimumab (TA187 [CD] and TA392 [UC]) Appraisal: 2002	For adults with severe active CD, or have not responded to conventional therapies	2002 (Initial); 2010 (updated)	• 2002: No • 2010: Yes, updated appraisal notes additional data on dose escalation from observational studies, used for treatment costs (unknown study design)
	For previously treated moderately to severely active UC	• 2012 (Initial) • 2015 (Additional evidence)	• 2012: No • 2015: Yes, an interim analysis of a prospective study looking at QoL and HRU, to understand HRU rates
Golimumab (TA392)	For previously treated moderately to severely active UC	2015	No
Infliximab (TA187 [CD] and TA392 [UC])	For adults with severe active CD, or have not responded to conventional therapies	• 2002 (Initial) • 2010 (Additional evidence)	• 2002: No • 2010: Yes, updated appraisal notes additional data on dose escalation from observational studies, used for treatment costs (unknown study design)
	For acute exacerbations (2008) For previously treated moderately to severely active UC (2015)	• 2008 (Initial) • 2015 (Additional evidence)	• 2008: No • 2015: No
Risankizumab (TA888 [CD] and TA998 [UC])	For previously treated moderately to severely active CD	2023	No
	For moderately to severely active UC	2024	Yes, this was a cost comparison submission that had RWE informing data for the comparison treatment in cost effectiveness models (unknown study design)
Ustekinumab (TA456 [CD] and TA633 [UC])	For previously treated moderately to severely active	2017	No
	For moderately to severely active UC	2020	No
Guselkumab	For moderately to severely active CD	Has not occurred yet	NA
	For moderately to severely active UC	Has not occurred yet	NA
Mirikizumab (TA925)	For moderately to severely active CD	Has not occurred yet	NA
	For moderately to severely active UC	2023	No
Vedolizumab (TA352 [CD] and TA342 [UC])	For moderately to severely active CD after prior therapy	2015	No
	For moderately to severely active UC	2015	No
JAK inhibitors			
Tofacitinib (TA547)	For moderately to severely active UC	2018	No
Filgotinib (TA792)	For treating moderately to severely active UC	2022	No
Upadacitinib (TA905 [CD] and TA856 [UC])	For previously treated moderately to severely active CD	2023	No
	For treating moderately to severely active UC	2023	No

Abbreviations: CD = Crohn's disease; HRU = health resource utilization JAK = Janus kinase; NA = not applicable; NICE = National Institute for Health and Care Excellence; QoL = quality of life; RWE = real-world evidence; TA = technology appraisal; UC = ulcerative colitis

- Sixteen advanced therapy single technology appraisals from 2015–2025 were reviewed.
- RWE was only included in one initial submission, a cost-comparison submission (for a UC therapy).
  - RWE for the comparator informed dose escalation proportions; no RWD used for the treatment appraised.
- RWE was also used in updated submissions for two treatments for CD to help understand treatment costs

Figure 2. Our considerations for integrating RWE into HTA submissions (initial or updated after initial reimbursement)

Abbreviations: NICE = National Institute for Health and Care Excellence; QoL = quality of life; RWE = real-world evidence

Conclusions

- **RWE is under-utilized in HTA submissions to NICE for advanced therapies for IBD.**
- **Future submissions should consider knowledge gaps that could most benefit from RWE integration, including modeling assumptions.**

References

1. Ingrasciotta Y, et al. *Pharmacol Res*. 2024 Feb;200:107074.
2. Vieujean S. et al. *Lancet Healthy Longev*. 2022 May;3(5):e356-e366.
3. NICE. NICE real-world evidence framework. 2022. Accessed October 19, 2023; https://www.nice.org.uk/corporate/ecd9/chapter/overview.

Disclosures

NRB and **JSS** are employees of PPD™ Observational Studies, Thermo Fisher Scientific. **GM** and **RM** are employees of PPD Evidera™ Health Economics and Market Access, Thermo Fisher Scientific. **DC** is an employee of PPD clinical research business of Thermo Fisher Scientific. Funding for this poster was provided by Thermo Fisher Scientific.

Acknowledgments

Editorial and graphic design support were provided by Michael Grossi and Shani Berger of Thermo Fisher Scientific.

Presented at the GetReal Conference 2025 | 1–2 October 2025 | Utrecht, The Netherlands

Use of Real-world Evidence in Advice Reports Assessing Reimbursement in The Netherlands from 2023–2025

ABS-20

Janneke Luijken[1], Ivana Sestak[2]

[1] Thermo Fisher Scientific, The Randstad, Netherlands; [2]Thermo Fisher Scientific, Milan, Italy

Background

- In 2025, the EMA released a reflection paper on the generation of real-world evidence (RWE).
- The reflection paper describes how RWE can add to existing evidence from interventional studies, providing data on effectiveness and safety that is representative of clinical practice.[1]
- The guideline published by Zorginstituut Nederland (ZIN) describes that RWE from observational studies can be included in economic evaluations and contribute larger external validity than interventional studies.[2]

Objectives

- **This study aimed to assess the use of RWE in submissions that are reviewed by ZIN for reimbursement advice.**
- **The goal was to create an overview and look at trends of RWE use and reimbursement recommendations.**

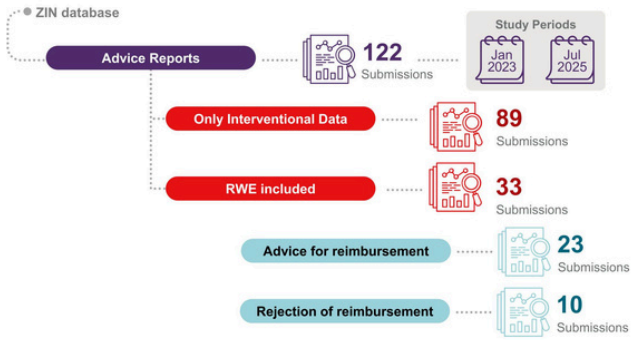
Methods

- A targeted review of advice reports published between January 2023 and July 2025 that evaluated new treatments for reimbursement in the basic insurance package was conducted in the ZIN database.[3]
- The reports included the submission with evidence provided by the market authorization holder (MAH).
- In the Netherlands, the basic insurance package is mandatory for all adults and covers all essential medical care.
- Relevant data was extracted from the submissions in the advice reports, including reimbursement recommendation decisions, type of RWE studies, and reasons for rejection. This data was summarized into a standardized data extraction form.
- The data are summarized with descriptive statistics, by presenting the number and proportion of advice reports in different categories.

Results

- A total of 122 advice reports were identified from the ZIN database.
- **Figure 1** presents a high-level overview of the results of this targeted review.
- Of the 33 submissions that included RWE, 14 (42.4%) included data from retrospective studies, eight (24.2%) data from prospective studies, and six (18.2%) used real-world data as external control arms; for the remaining five submissions, type of RWE was not specified or data from databases/electronic health records were used.

Figure 1. High-level Overview of Targeted Review Results

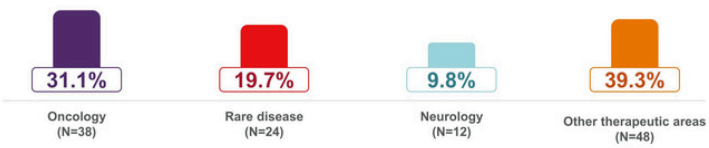


Abbreviations: RWE = real-world evidence; ZIN = Zorginstituut Nederland

Therapeutic Area

- The most common therapeutic area was oncology, with 38 (31.1%) submissions.
- **Figure 2** presents an overview of the most common therapeutic areas.

Figure 2. Therapeutic Areas Assessed in Submissions

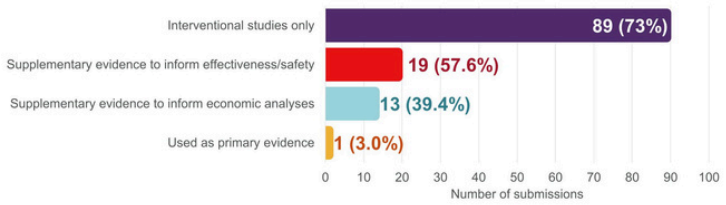


Results (cont.)

Type of Evidence

- Most submissions included only data from interventional studies (N=89, 73.0%).
- The remaining 33 submissions (27.0%) included some type of RWE (**Figure 3**).
  - Only one (3.0%) submission used RWE as the primary evidence.
  - Nineteen (57.6%) submissions used RWE as supplementary evidence to inform effectiveness/safety.
  - Thirteen (39.4%) submissions used RWE as supplementary evidence to inform economic analyses.

Figure 3. Type of Evidence in all Submissions, n (%)



Therapeutic Areas in Relation to RWE Use

- Half (50%) of the oncology submissions incorporated RWE in the submissions.
- In contrast, only 8.3% of neurology submissions included RWE.
- The five therapeutic areas with the highest percentage of RWE use in submissions is shown in **Table 1**.

Table 1. Top 5 Therapeutic Areas with RWE Use in Submissions

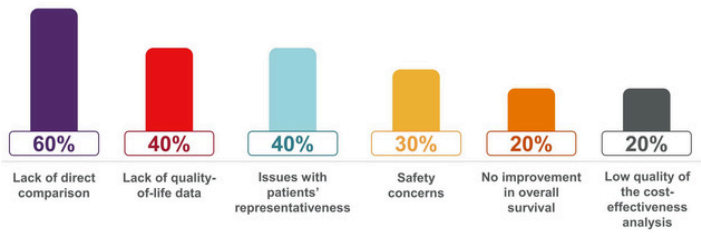
Therapeutic area (submissions)	Percentage of submissions including RWE
Oncology (n=38)	50.0%
Nephrology (n=2)	50.0%
Cardiology (n=5)	40.0%
Rare disease (n=24)	25.0%
Hematology (n=4)	25.0%

Abbreviation: RWE = real-world evidence

Reasons for Withholding Reimbursement Advice in Submissions with RWE

- The reasons for withholding reimbursement advise were summarized. Out of the 10 submissions that included some type of RWE but did not receive a reimbursement recommendation, "*lack of a direct comparison*" was listed in six advice reports.
- **Figure 4** presents the different reasons listed in the 10 advice reports.

Figure 4. Reasons for Rejection of Reimbursement Advice



Conclusions

- **Use of RWE in submissions in the Netherlands is relatively low, with 27% of submissions including RWE.**
- **Irrespective of type of data included (interventional or RWE), studies that include a comparator and quality of life assessments, with an unbiased representation to the general target population, are needed to support the evidence portfolio of new treatments.**

References

1. EMA. Reflection paper on use of real-world data in non-interventional studies to generate real-world evidence for regulatory purposes. 2025. Accessed September 16, 2025. https://www.ema.europa.eu/en/documents/other/reflection-paper-use-real-world-data-non-interventional-studies-generate-real-world-evidence-regulatory-purposes_en.pdf
2. ZIN. Richtlijn voor het uitvoeren van economische evaluaties in de gezondheidzorg. 2024. Accessed September 16, 2025. https://www.zorginstituutnederland.nl/site/binaries/site-content/collections/documents/2024/01/16/richtlijn-voor-het-uitvoeren-van-economische-evaluaties-in-de-gezondheidszorg/Richtlijn+voor+het+uitvoeren+van+economische+evaluaties+in+de+gezondheidszorg+%28versie+2024%29.pdf
3. ZIN. Publicaties. 2025. Accessed September 16, 2025. https://www.zorginstituutnederland.nl/documenten

Disclosures

JL and **IS** are employees of PPD™ Observational Studies, Thermo Fisher Scientific. Funding provided by Thermo Fisher Scientific.

Acknowledgements

Editorial and graphic design support were provided by Michael Grossi, Shani Berger, and Kawthar Nakayima of Thermo Fisher Scientific.

Is There a Lack of Dietary Data Collection in Real-world IBD Studies? Literature Review and Future Considerations

ABS-21

Neil R. Brett,[1] Garthiga Manickam,[1] Marielle Bassel,[1] Daniela Castano,[2] John S. Sampalis[1]

[1]Thermo Fisher Scientific, Montreal, Canada; [2]Thermo Fisher Scientific, Waltham, MA, USA

Background

- Leading gastroenterology associations (e.g., American Gastroenterological Association [AGA] and the European Crohn's and Colitis Organization [ECCO]), have recent guidelines that highlight evidence showing that healthy dietary patterns in patients with inflammatory bowel disease (IBD) can help manage their disease, reduce progression, or induce remission.[1,2]
- However, due to cost, data availability, and collection burden, it is unclear how often diet is ascertained in observational IBD studies.
- This gap assessment will help guide recommendations for integrating dietary data into real-world evidence and applying to clinical practice.

Objective

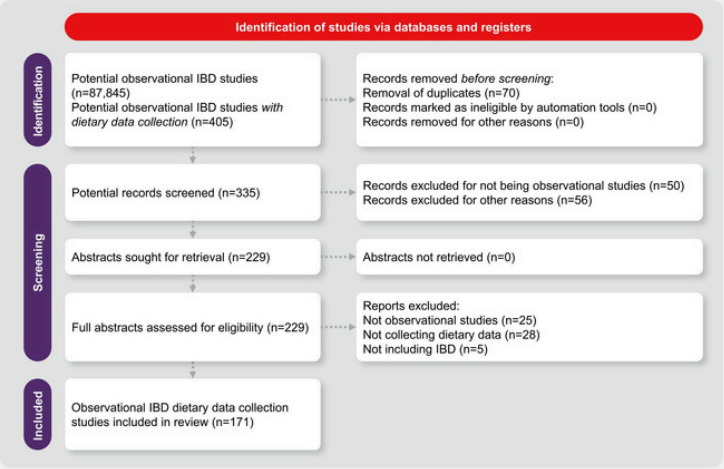
- **The objective was to describe dietary data collection methods and purposes in observational IBD studies and provide considerations to enhance future research.**

Methods

- A targeted literature review was conducted to identify observational IBD studies that collected dietary data. Searches were performed for IBD or ulcerative colitis (UC) or Crohn's disease (CD) and nutrition or diet or dietary or food or diet record or diet questionnaire.
- Records were retrieved from the databases and de-duplicated. Two individuals independently reviewed the remaining abstracts and full-text studies to ensure they met the search criteria and extracted key information, such as population, disease, data collection methods, and study outcomes.

Results

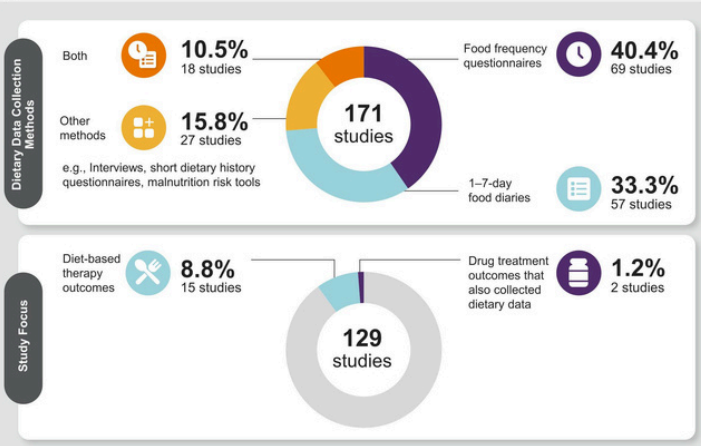
Figure 1. PRISMA Flow Diagram of Abstract Screening



Abbreviations: IBD = inflammatory bowel disease; PRISMA = Preferred Reporting Items for Systematic Reviews and Meta-analyses

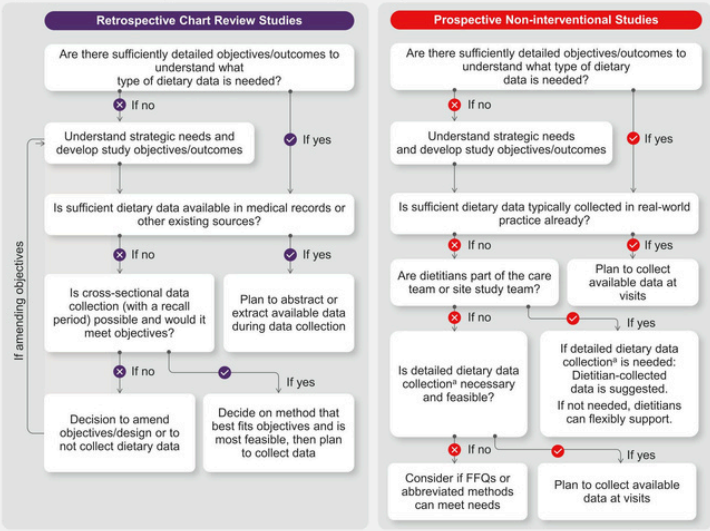
- Of the 171 studies included, 40 (23.4%) studies were in CD, 24 (14.0%) were in UC, and 107 (62.6%) included both. One-hundred and nineteen studies (69.6%) focused on adults, while the remainder included children or both age groups.

Figure 2. Data Collection Methods



Key Finding (Figure 2): Despite known links between diet and IBD outcomes, only 1.2% of reviewed studies collected dietary data to support assessment of drug treatment effectiveness.

Figure 3. Flow Diagrams For Selecting Dietary Data Collection Methods



Abbreviations: FFQ = Food Frequency Questionnaire. [a]Detailed dietary data collection could include 24-hour recalls or food diaries.

Figure 4. Considerations For Selecting Dietary Data Collection Methods

Alignment of Dietary Data Collection to Objectives/Outcomes	• To assess general dietary patterns (e.g., diet frequency, dietary quality) and association with outcomes – Use brief tools such as FFQs or short questionnaires	• To associate specific nutrients with disease outcomes – Use detailed methods (e.g., food diaries, multiple 24-hour recalls)
Leveraging Existing Data Sources (When Possible)	• Patients may have some dietary data in clinical records: – Inpatients or those receiving enteral/parenteral nutrition – Pediatric or adult patients on structured dietary interventions – Patients with diagnosed malnutrition or under care of dietitians • Feasibility checks needed to determine level of detail available	
Cross-sectional Data Collection in Retrospective Studies	• Use validated FFQs: – Low burden, capture usual intake over the past 30 days to 1 year – Can be administered electronically or on paper	• Limitations: – Potential recall bias – FFQ recall period may not align with entire follow-up period – Requires patient consent
Prospective Data Collection	• FFQs or brief dietary surveys at multiple timepoints: – Capture general intake patterns	• Food diaries (3–7 days) or 24-hour recalls: – Best for tracking detailed macro/micronutrient intake – Require more participant effort • Involving dietitians is strongly recommended: – Improves data quality – Supports flexible recall methods (in-person/phone) – Reduces participant burden
Technology Considerations	• Nutrition-specialized digital platforms: – Guided input (e.g., prompts, pictures) reduces missing data – Enables real-time data analysis and monitoring	• Emerging AI tools (e.g., food image recognition): – Improve data quality through visual verification – Allows proxy input (e.g., caregivers, parents) – May reduce patient burden

Abbreviations: AI = artificial intelligence; FFQ = Food Frequency Questionnaire

Conclusions

- **Dietary data collection methods varied among observational IBD studies.**
- **Few studies collected dietary data when evaluating drug treatment outcomes.**
- **Future study designs should align data collection methods with research objectives/ population and balance participant burden with data quality**
- **Many real-world limitations can be minimized through planning—including early input from clinical, nutrition, and operational experts during study design and start-up.**

References

1. Hashash JG, et al. *Gastroenterology*. 2024 Mar;166(3):521-532.
2. Vaios Svolos V, et al. *J Crohns Colitis*. Published online July 12, 2025.

Disclosures

NRB and **NDM** are employees of PPD™ Observational Studies, Thermo Fisher Scientific. **GM** is an employee of PPD Evidera™ Health Economics & Market Access. **DC** is employed by PPD clinical research business of Thermo Fisher Scientific. Funding provided by Thermo Fisher Scientific.

Editorial and graphic design support were provided by Michael Grossi and Richard Leason of Thermo Fisher Scientific.

Presented at the GetReal Conference 2025 | 1–2 October 2025 | Utrecht, The Netherlands

Are Cross-sectional Cohorts an Efficient Alternative to Prospective Cohort Design in Real-world Studies?

ABS-22

Neil R. Brett,[1] Marielle Bassel,[1] Vincent McCarty,[2] John Sampalis[1,2]

[1]Thermo Fisher Scientific; Montreal, Canada; [2]McGill University, Montreal, Canada

Background

- Longitudinal prospective cohort (LPC) study design may be considered the gold standard for real-world studies of change over time in clinical or patient-reported outcomes.
- However, these designs are sometimes non-feasible or cost prohibitive due to long study durations, patient burden, regulatory requirements, and/or logistical/operational complexities.[1,2]
  - Cross-sectional cohort (CSC) design, in which subjects are assessed once but at varying timepoints of their disease journey, may mitigate many of the design challenges posed by LPC studies.
  - However, it is unknown how well CSC studies can replicate the outcomes demonstrated in LPC studies.
- Thus, research is needed comparing CSC with LPC study designs.

Objective

- **The objective was to compare CSC study design with LPC study design in a cohort of patients with rheumatoid arthritis (RA).**

Methods

- The study utilized real-world data from 1,716 patients with RA receiving either of two treatments.
- Clinical Disease Activity Index (CDAI) was assessed at 0, 3, 6, 9, and 12 months in the LPC study.
- To simulate the CSC study, a single timepoint was randomly selected as the hypothetical cross-sectional time of assessment.
- Univariate and repeated measures general linear models were used to analyze the data.
  - Between-treatment differences were adjusted in both models for age, gender, duration of disease, prior treatments, and markers of inflammation.

Results

Table 1. Baseline characteristics

	Treatment					
	A		B		C	
Age: Mean (SD) years	59.62	(11.46)	60.29	(11.61)	59.94	(11.54)
Gender N(%)						
Female	6,255	73.90%	617	-74.40%	1272	83.10%
Male	231	(26.1%)	213	(25.7%)	444	(25.9%)
Disease Duration Mean (SD) years	7.18	(8.76)	8.02	(9.18)	7.59	(8.97)
CRP Mean (SD) mg/L	2.10	(3.55)	2.01	(3.05)	2.06	(3.32)
ESR Mean (SD) mm/hr	22.38	(17.13)	23.32	(16.63)	22.84	(16.89)
RF+ N(%)	484	(54.6%)	494	(59.5%)	978	(57.0%)
ACCP+ N(%)	284	(32.1%)	280	(33.7%)	564	(32.9%)
Prior Treatments						
Corticosteroids N(%)	211	(23.8%)	185	(22.3%)	396	(23.1%)
NSAIDS N(%)	268	(30.2%)	230	(27.7%)	498	(29.0%)
HCQ N(%)	354	(40.0%)	336	(40.5%)	690	(40.2%)
Sulfasalazine N(%)	142	(16.0%)	140	(16.9%)	282	(16.4%)
Leflunomide N(%)	112	(12.6%)	116	(14.0%)	228	(13.3%)
Gold N(%)	7	(0.8%)	5	(0.6%)	12	(0.7%)
Biologic DMARD N(%)	359	(40.5%)	331	(39.9%)	690	(40.2%)

Abbreviations: ACCP+ = anti-cyclic citrullinated peptide-positive; CRP = C-reactive protein; DMARD = disease-modifying anti-rheumatic drug; ESR = erythrocyte sedimentation rate; HCQ = hydroxychloroquine; NSAID = nonsteroidal anti-inflammatory drug; RF+ = rheumatoid factor-positive

Table 2. Estimated least squared mean CDAI

		Cross Sectional Cohort				Longitudinal Prospective Cohort			
				95% CI				95% CI	
Treatment	Month	Mean	SE	Lower Bound	Upper Bound	Mean	SE	Lower Bound	Upper Bound
A	0	29.36	0.971	27.459	31.270	27.49	1.975	23.619	31.365
	3	17.88	0.747	16.412	19.343	17.23	2.035	13.237	21.221
	6	15.55	0.729	14.125	16.983	15.86	2.118	11.704	20.012
	9	12.45	0.670	11.131	13.761	13.13	1.700	9.799	16.469
	12	10.52	1.003	8.553	12.488	12.95	1.861	9.297	16.597
B	0	31.10	0.997	29.140	33.051	27.31	1.976	23.437	31.190
	3	15.97	0.730	14.535	17.398	17.76	2.037	13.763	21.754
	6	16.17	0.732	14.735	17.606	15.24	2.120	11.086	19.401
	9	11.85	0.751	10.377	13.324	11.77	1.702	8.435	15.111
	12	12.24	1.039	10.207	14.281	12.15	1.863	8.494	15.800
Total	0	30.23	0.696	28.865	31.595	29.79	0.245	29.309	30.270
	3	16.92	0.522	15.898	17.947	16.20	0.253	15.702	16.695
	6	15.86	0.517	14.848	16.876	15.04	0.263	14.520	15.550
	9	12.15	0.503	11.161	13.135	12.09	0.211	11.679	12.505
	12	11.38	0.722	9.966	12.799	12.06	0.230	11.605	12.509

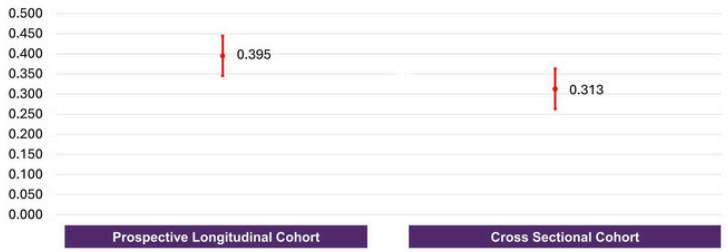
Figure 1. Estimated least squared mean CDAI

Table 3. Covariates

	P-value	
Variable	Cross-sectional Cohort	Longitudinal Prospective Cohort
Treatment	0.559	0.806
Month	0.001	0.001
Treatment * Month	0.136	0.655
Age	0.356	0.501
Gender	0.006	0.025
Disease Duration	0.051	0.025
Baseline CRP	0.061	0.003
Baseline ESR	0.764	0.734
RF	0.000	0.000
ACCP	0.055	0.001
Prior Treatments		
MTX	0.000	0.000
Corticosteroids	0.591	0.477
NSAIDS	0.001	0.000
HCQ	0.008	0.079
Sulfasalazine	0.451	0.040
Leflunomide	0.935	0.161
Gold	0.281	0.957
Biologic	0.000	0.000

Abbreviations: ACCP+ = anti-cyclic citrullinated peptide; CRP = C-reactive protein; ESR = erythrocyte sedimentation rate; HCQ = hydroxychloroquine; MTX = methotrexate; NSAID = nonsteroidal anti-inflammatory drug; RF = rheumatoid factor

Figure 3. Between-group differences



Discussion

- The results of this study showed that the CSC analysis (using general linear models) emulated the results of an LPC analysis (using a repeated measures mixed effects model).
  - Between-group differences with the two analyses were similar.
  - The two methods showed comparable statistical significance for most covariates included in the models.

Limitations

- CSC was simulated from a longitudinal study by selecting random visit numbers at different disease intervals.
- Because timing of assessment in real-world studies may be due to multiple known and unknown factors (i.e., may not be random), assessments of differential attrition and visit patterns would be required to ensure that the CSC model can be used. Adjustments for informed censoring may be necessary.
- This study was based on RA, a progressive chronic condition, where disease severity increases with time.
  - Thus, the results may not apply to diseases where severity does not follow a linear trend.
- While the example was derived from RA, it is possible that the results could be generalized to other paradigms where longitudinal assessments of disease are needed.

Conclusions

- **Cross-sectional analyses could be a valid replacement for longitudinal prospective analyses in some instances.**
- **A CSC approach offer advantages over the longitudinal studies with respect to duration and costs.**
- **Careful considerations must be given to potential issues of bias and suitability of the therapeutic area for utilization of the CSC approach.**

References

1. Alarkawi D, et al. *JBMR Plus*. 2018 Apr 30;2(4):187-194.
2. Saag KG. Clinical Research Methods in the Rheumatic Diseases. In: Firestein GS, Budd RC, Gabriel SE, et al, eds. Kelley's Textbook of Rheumatology. 9th ed. *Elsevier*; 2013:495–510.

Disclosures

NRB, **MB** and **JS** are employees of PPD™ Observational Studies, Thermo Fisher Scientific. Funding for this poster was provided by Thermo Fisher Scientific.

Acknowledgments

Editorial and graphic design support were provided by Michael Grossi and Shani Berger of Thermo Fisher Scientific.

Data Reliability in Retrospective Chart Review Studies: Results and Considerations from a Novel Data Review Methodology

ABS-23

Neil R. Brett,[1] Luke Safarcyk,[2] Elizabeth Donahue,[2] Marielle Bassel[1]

[1]Thermo Fisher Scientific, Montreal, Canada; [2]Thermo Fisher Scientific, Waltham, MA, USA

Background

- Retrospective historical cohort studies (chart review studies) are generally efficient designs to generate longitudinal observational follow-up data.[1]
- For data review and abstraction verification in clinical studies, source data verification (SDV) may be thought of as the gold-standard.[2]
  - However, the potential use of SDV in chart review studies may face challenges in implementation, site interest, data protection, and timelines.
  - Thus, chart review studies generally use remote data review (without source documents); however, this method can result in data reliability/accuracy issues for studies.
- There is a need for data review methodology that is realistic for efficient implementation in the chart review context, yet still with additional rigor compared with traditional remote data review.

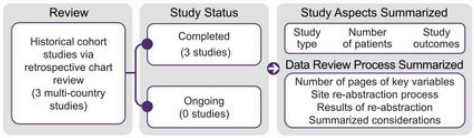
Objective

- **The objective was to describe an alternative method for data quality review and highlight key considerations.**

Methods

- We performed a review of the three historical cohort studies via retrospective chart review.
- The summary of study status and key evaluations conducted are presented in the **Figure 1**.

Figure 1. Three historical cohort studies via retrospective chart review

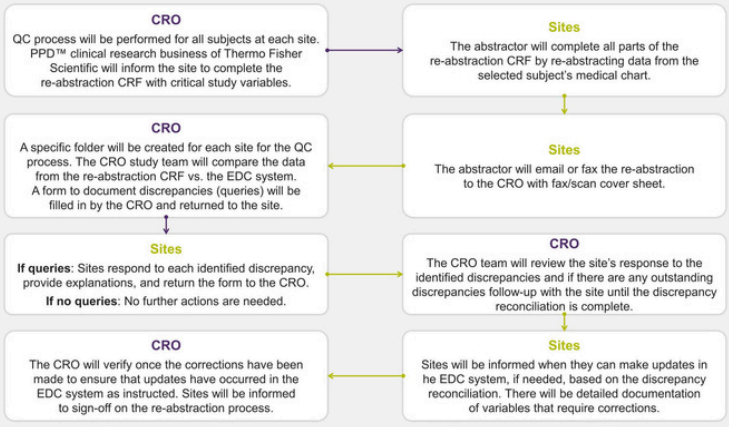


Results

Table 1. Overview of Study Types, Populations and Key Outcomes

Site-based or De-centralized	Study Population and N	Therapeutic Area	Type of Chart Review Study	Types of Data Collected
Site-based	Adults, n=227	Gastroenterology	Natural History Study	Medical history, treatment patterns, clinical outcomes
Site-based	Pediatrics, n=30	Rare Disease	Label Expansion	Patient demographics, medical history, initiation of treatment, clinical outcomes
Site-based	Adults, n=18	Immunology	Drug Utilization	Patient demographics, medical history, treatment patterns including adherence, laboratory tests

Figure 2. Data Review Process Overview of Repeat Abstraction of Key Variables



Abbreviations: CRF = case report form; CRO = contract research organization; EDC = electronic data capture; QC = quality control

Figure 3. Details and Discrepancies for Re-abstraction of Variables

Number of pages/variables of repeat abstraction per study:

- **Study 1:** **1 page** including **13 variables**
- **Study 2:** **3 pages** including **22 variables**
- **Study 3:** **7 pages** including **61 variables**

Number of patients/sites per study included in re-abstraction:

- **Study 1:** **7 sites, re-abstraction for 1 patient per site** (7 total patients)
- **Study 2:** **17 sites, re-abstraction for all patients** (30 total patients)
- **Study 3:** **5 sites, re-abstraction for 1 patient per site** (5 total patients)

Re-abstraction discrepancies:

- **Study 1:** **4 patients** with **0 discrepancies**, **1 patient** with **1%–9% discrepancies** and **2 patients** with **30%– 40% discrepancies**
- **Study 2:** **26 patients** with **10%– 39% discrepancies, 4 patients** with **>40% discrepancies**
- **Study 3:** **0 discrepancies for all patients**

Figure 4. Considerations for Repeat Abstraction of Key Variables Data Review

- Study types that may be suitable for this methodology: Retrospective studies where on-site monitoring or source data verification may not occur.

Conclusions

- **The data review methodology of repeat abstraction of key variables demonstrates the importance of additional data QC in chart review studies, which enhances the reliability and accuracy of data abstraction, leading to more robust study outcomes.**
- **This cost-efficient methodology should be customized to each study based on study design, outcomes, timelines, and other relevant factors.**
- **Studies looking to implement this data review methodology should tailor it to be study-specific by carefully considering data abstraction complexity, study timelines, site burden, and purpose of the study.**

References

1. Saag KG, et al. Clinical Research Methods in the Rheumatic Diseases. In: Firestein GS, Budd RC, Gabriel SE, et al., eds. Kelley and Firestein's Textbook of Rheumatology. 10th ed. Vol 1. *Elsevier*; 2017:475–485.
2. Andersen JR, et al. *Br J Clin Pharmacol*. 2015 Apr;79(4):660-8.
3. Hamidi M, et al. *J Clin Transl Sci*. 2024 May 21;8(1):e101.

Disclosures

All authors are employees of PPD™ Observational Studies, Thermo Fisher Scientific. Funding provided by Thermo Fisher Scientific.

Acknowledgments

Editorial and graphic design support were provided by Michael Grossi and Richard Leason of Thermo Fisher Scientific.

Presented at the GetReal Conference 2025 | 1–2 October 2025 | Utrecht, The Netherlands

Cost-effectiveness of Radiotherapy in Uterine Serous Carcinoma (USC): A Real-world Study

ABS-25

Elefterios (Pierre) Samartzis,[1] Lucy Gilbert,[2] Vincent McCarty,[2] Neil Brett,[3] Marielle Bassel,[3] John Sampalis[2,3]

[1]University Hospital Zürich, Zürich, Switzerland; [2]Thermo Fisher Scientific, Montreal, Canada; [2]McGill University, Montreal, Canada

Background

- Approximately 10% of all endometrial cancers (ECs) are uterine serous (USC).[1]
- Distribution of stage at diagnosis is predominantly stage I (40%), followed by stage II (30%), stage III (20%), and stage IV (10%).[2]
- However, due to its aggressive nature, USC EC is associated with high recurrence and poor prognosis, and it accounts for 40% of EC-related deaths.[2]
- Treatment of EC may consist of multiple components.
  - Adjuvant chemotherapy (C) with carboplatin + paclitaxel is always recommended.
  - Surgery recommendations depend on diagnosis (e.g., total hysterectomy, bilateral salpingo-oophorectomy, pelvic and para-aortic lymph node dissection, omental biopsy and pelvic washings for staging and detection of spread cancer).
  - The benefit of adjuvant radiotherapy (vaginal brachytherapy or external beam radiotherapy [EBRT]) is currently unclear.
- Since the potential benefit and optimal timing of adjuvant radiotherapy is unclear additional real-world evidence is needed to understand its effectiveness.

Objective

- **To describe the cost-effectiveness of adjuvant radiotherapy in USC patients treated in the real-world setting.**

Methods

- Retrospective observational cohort study
- Patients with USC receiving:
  - C or C + adjuvant external beam radiotherapy (EBRT)
- All patients were treated at the McGill University Health Center (MUHC) between 2008 and 2023.
  Patient treatment characteristics and outcomes were ascertained from the MUHC Electronic Health Records and the MUHC Gyno-Oncology Database.
- Costs for radiotherapy were assumed to be 2025 average costs in the US.
- Incremental cost-effectiveness ratios (ICERs) were estimated per year of progression-free (PFS) and overall survival (OS) benefit.

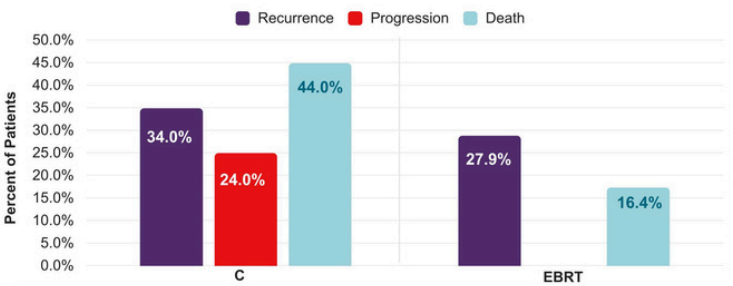
Results

Table 1. Patient characteristics and outcomes

		Treatment Group				*P*-value
		C (n=50)		C+EBRT (n=61)		
Mean (SD) age, years		67.30	(9.07)	69.40	(8.11)	0.219
Stage I		13	26.0%	29	47.5%	<0.001
Stage II		0	0.0%	13	21.3%	
Stage III		13	26.0%	16	26.2%	
Stage IV		24	48.0%	1	1.6%	
Preop albumin	< 3.5	6	12.0%	1	1.6%	0.028
Cytology	Malignant	23	46.0%	9	14.8%	<0.001
MMR Status	Deficient	1	2.0%	7	11.5%	0.046

Abbreviations: C = chemotherapy; EBRT = external beam radiotherapy; MMR = mismatch repair status

Figure 1. Proportion of patients with clinical outcomes



	Treatment Group				*P*-value
	C (n=50)		C+EBRT (n=61)		
Recurrence	17	34.0%	17	27.9%	0.560
Progression	12	24.0%	0	0.0%	<0.001
Death	22	44.0%	10	16.4%	0.005

Abbreviations: C = chemotherapy; EBRT = external beam radiotherapy

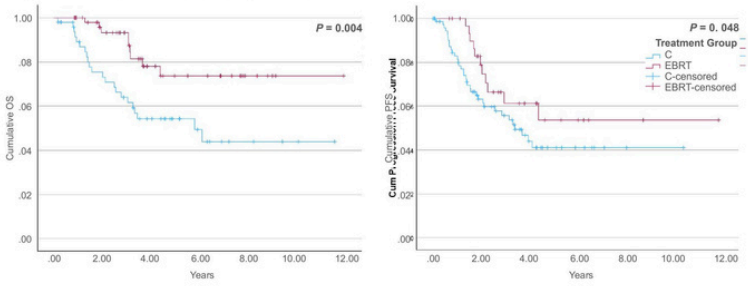
References

1. Bogani G, et al. *Gynecol Oncol. 2021;162(1):226-34.*
2. Ball A, et al. *Journal of Obstetrics and Gynaecology Canada. 2014;36(12):1085-92.*

Results (cont.)

Cumulative OS and PFS

Figure 2. Cumulative OS and PFS



Abbreviations: C = chemotherapy; EBRT = external beam radiotherapy; PFS = progression-free survival

Regression and Cost-effectiveness

Table 2. Cox regression analyses understanding how covariates associate with clinical outcomes

	OS				PFS			
			95.0% CI for HR				95.0% CI for HR	
	P-value	HR	Lower	Upper	*P*-value	HR	Lower	Upper
EBRT vs C	0.242	0.571	0.223	1.459	0.141	0.584	0.285	1.196
Stage (III–IV) vs (I–II)	0.040	2.726	1.046	7.109	0.004	2.915	1.403	6.053
Pre-Op Albumin <= 3.5	0.685	1.362	0.306	6.056	0.926	0.934	0.218	4.004
MMR Deficient	0.978	0.000	0.000	—	0.389	0.531	0.126	2.242

MMR = Mismatch Repair Abbreviations: C = chemotherapy; EBRT = external beam radiotherapy; HR = hazard ratio; OS = overall survival; PFS = progression free survival

Table 3. Describing cost-effectiveness of treatments

Treatment	N	Cost per EBRT	Total Cost for EBRT	Total Survival Years	Total PFS Years
EBRT	100	$27,812.00	$2,781,200.00	998.7	745.8
C	100	$0.00	$0.00	681.9	536.4
Difference (EBRT-C)				316.8	209.4
Incremental cost per year				$8,779.04	$13,281.76

Abbreviations: C = chemotherapy; EBRT = external beam radiotherapy; PFS = progression free survival

Discussion

- The results of this study have shown that adjuvant radiation therapy provides clinically important benefits in terms of OS and PFS for patients with USC EC.
- Adjuvant radiation therapy had an acceptable ICER for both OS and PFS.

Limitations

- This was a single-site study conducted in a tertiary center that is highly specialized in the treatment of gynecological cancers:
  - Therefore, the study population may not be representative of the general patient population with USC EC.
- This study was conducted in Canada, where a universal, publicly funded healthcare system is in effect.
  - Results may be different in non-public/universal healthcare systems, where access to care may be a barrier to receiving radiation therapy.
- The study was conducted prior to increased use of targeted and immunotherapy.
  - Thus, results may vary among inpatients treated with advanced therapies.

Conclusions

- **Adjuvant radiation therapy (EBRT) may be beneficial and cost-effective in a universal healthcare system.**
- **Further studies are needed to confirm these results in other health care systems.**
- **Real-world studies are required to conduct cost-effectiveness assessments of treatments.**
  - **Evidence from these studies will drive decisions regarding optimal use of high-cost treatments for rare diseases and cancers.**

Disclosures

NRB, **MB**, and **JS** are employees of PPD™ Observational Studies, Thermo Fisher Scientific. Funding for this poster was provided by Thermo Fisher Scientific.

Acknowledgements

Editorial and graphic design support were provided by Caroline Cole and Kawthar Nakayima of Thermo Fisher Scientific.

A Real-world-derived Algorithm to Improve the Cost-effectiveness of External Beam Radiotherapy in Uterine Serous Carcinoma

ABS-26

Elefterios (Pierre) Samartzis,[1] Lucy Gilbert,[2] Vincent McCarty,[2] Neil R. Brett,[3] Marielle Bassel,[3] John Sampalis[2,3]

[1]University Hospital Zürich, Zürich, Switzerland; [2]Thermo Fisher Scientific, Montreal, Canada; [3]McGill University, Montreal, Canada

Background

- Approximately 10% of all endometrial cancers (ECs) are uterine serous (USC)[1].
- Distribution of stage at diagnosis is predominantly stage I (40%), followed by stage II (30%), stage III (20%), and stage IV (10%)[2].
- However, due to its aggressive nature, USC EC is associated with high recurrence and poor prognosis, and it accounts for 40% of EC-related deaths[2].
- Treatment of EC may consist of multiple components.
  - Adjuvant chemotherapy (C) with carboplatin + paclitaxel is always recommended.
  - Surgery recommendations depend on diagnosis (e.g., total hysterectomy, bilateral salpingo-oophorectomy, pelvic and para-aortic lymph node dissection, omental biopsy and pelvic washings for staging and detection of spread cancer).
  - The benefit of adjuvant radiotherapy (vaginal brachytherapy or external beam radiotherapy [EBRT]) is currently unclear.
- Since the potential benefit and optimal timing of adjuvant radiotherapy is unclear, understanding its effectiveness and identifying patients who would benefit from EBRT would optimize the utility of this treatment.

Objective

- **The objective was to develop and evaluate an empirical algorithm aimed at identifying patients with USC EC who would benefit from adjuvant radiotherapy.**

Methods

- Retrospective observational cohort study
- Patients with USC receiving:
  - C
  - C + adjuvant EBRT
- All patients were treated at the McGill University Health Center (MUHC) between 2008 and 2023.
- Patient treatment characteristics and outcomes were ascertained from the MUHC Electronic Health Records and the MUHC Gyno-Oncology Database.
- Multi-variate logistic regression was used to identify predictors of progression or recurrence among those treated with EBRT.
- A receiver operating curve analysis was used to identify three groups with high, moderate, and low potential benefit from EBRT.

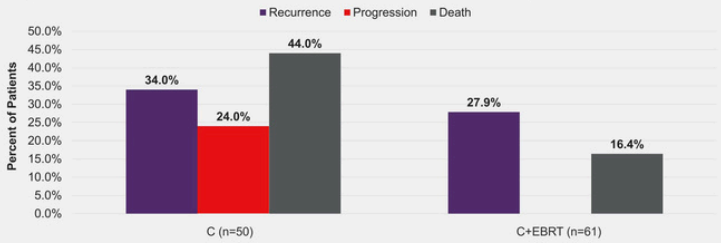
Results

Table 1. Patient characteristics

		Treatment Group				*P*-value
		C (n=50)		C+EBRT (n=61)		
Mean (SD) age, years		67.30	(9.07)	69.40	(8.11)	0.219
Stage I		13	26.0%	29	47.5%	
Stage II		0	0.0%	13	21.3%	<0.001
Stage III		13	26.0%	16	26.2%	
Stage IV		24	48.0%	1	1.6%	
PR+		27	54.0%	40	65.6%	0.028
Cytology	**Malignant**	23	46.0%	9	14.8%	<0.001

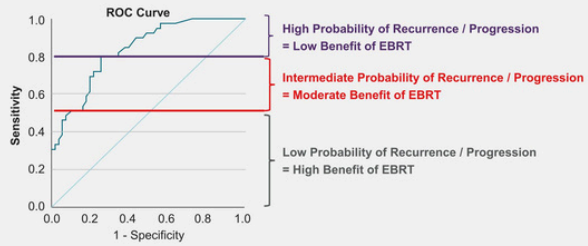
Abbreviations: C = chemotherapy; EBRT = external beam radiotherapy; PR = progesterone receptor

Figure 1. Proportion of patients with outcomes, stratified by treatment



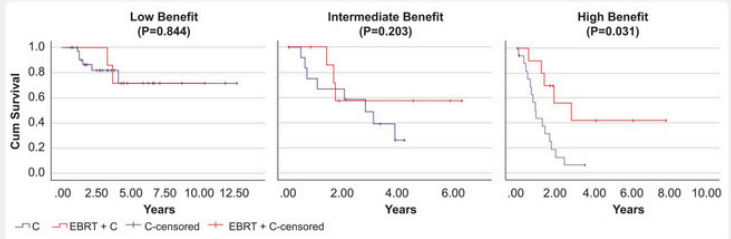
Abbreviations: C = chemotherapy; EBRT = external beam radiotherapy; PRISMA = Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Figure 2. Classification of EBRT benefit categories (*P*<0.001)



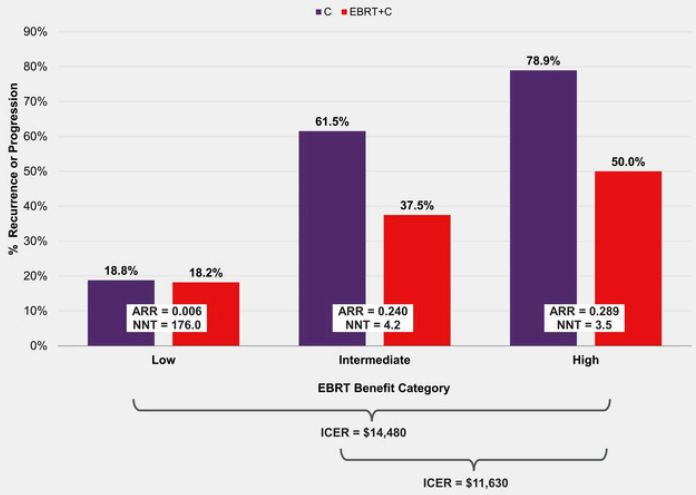
Abbreviations: EBRT = external beam radiotherapy; ROC = receiver operating characteristic

Figure 3. Progression-free survival by benefit categories



Abbreviations: C = chemotherapy; EBRT = external beam radiotherapy

Figure 4. Recurrence or progression by EBRT benefit category



Abbreviations: ARR = absolute risk reduction; EBRT = external beam radiotherapy; ICER = incremental cost effectiveness ratio / progression-free survival (PFS) year gained; NNT = numbers needed to treat

Discussion

- The real-world-derived algorithm for the selection of patients with USC ECs to be treated with EBRT has the potential to:
  - Improve the number needed to treat (NNT)
  - Reduce overall costs and improve cost-effectiveness
  - Prevent potentially non-beneficial EBRT in 33/61 (54.1%) of the patients
    - With associated impact on quality of life and treatment-related adverse events

Limitations

- This was a single-site study conducted in a tertiary center that is highly specialized in the treatment of gynecological cancers.
  - Therefore, the study population may not be representative of the general patient population with USC EC.
- This study was conducted in Canada, where a universal, publicly funded healthcare system is in effect.
  - Results may be different in non-public/universal healthcare systems, where access to care may be a barrier to receiving radiation therapy.
- The study was conducted prior to increased use of targeted and immunotherapy.
  - Thus, results may vary among inpatients treated with advanced therapies.

Conclusions

- **Adjuvant radiation therapy (EBRT) may be beneficial and cost–effective in a universal healthcare system.**
- **Using a real-word-derived algorithm to select patients who are most likely to benefit from EBRT may reduce overall costs.**

References

1. Bogani G, et al. Uterine serous carcinoma. Gynecol Oncol. 2021 Jul;162(1):226-234.
2. Ball A, et al. The importance of surgical staging in women with uterine serous carcinoma: experience in a single institution reveals a survival benefit. J Obstet Gynaecol Can. 2014 Dec;36(12):1085-1092.

Disclosures

NRB, MB, and **JS** are employees of PPD™ Observational Studies, Thermo Fisher Scientific. Funding for this poster was provided by Thermo Fisher Scientific.

Editorial and graphic design support were provided by Caroline Cole and Richard Leason of Thermo Fisher Scientific.

Presented at the GetReal Conference 2025 | 1–2 October 2025 | Utrecht, The Netherlands

Comparing Randomized Clinical Trials to Real-world Studies Evaluating the Effectiveness of a bDMARD in the Management of Crohn's Disease

ABS-27

Vincent McCarty[1], Paige Kostoulias[1], Neil R. Brett[2], Marielle Bassel[2], John Sampalis[1,2]

[1] McGill University, Montreal, Canada; [2] Thermo Fisher Scientific, Montreal, Canada

Background

- Recent studies in inflammatory bowel disease (IBD) highlight differences between randomized controlled trial (RCT) populations and populations in RWS, including age, comorbidities, disease severity, and other parameters that could affect treatment responses.[1,2]
  - Because of this, real-world studies (RWS) are becoming increasingly impactful in supporting evidence of treatment effectiveness and safety due to better generalizability to the target population.
- There are currently multiple biologic disease-modifying antirheumatic drugs (bDMARDs) approved for Crohn's disease (CD). However, it is not known if RCTs of bDMARDs have similar patient populations and outcomes to RWS of these treatments.

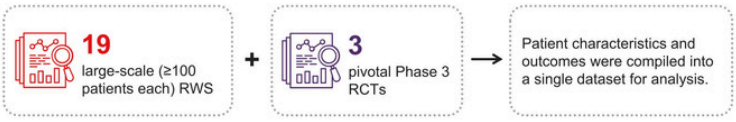
Objective

- **To compare patient profiles and outcomes between RCTs and RWS in the evaluation of a bDMARD treatment for CD.**

Methods

Figure 1. Study design

Data on patients treated for CD with the same bDMARD (identified via PubMed database search) were extracted from:



Abbreviations: CD = Crohn's disease; bDMARD = biologic disease-modifying antirheumatic Drug; RCT = randomized controlled trial; RWS = real-world studies

Statistics

- Categorical variables (sex, prior anti-tumor necrosis factor [TNF] exposure, immunosuppressant and corticosteroid use, Montreal Classification, clinical remission rates) were compared between RWS and RCT cohorts using two-sample z-tests for proportions. Statistical significance for categorical comparisons was defined as p<0.05.
- Continuous variables (age, disease duration, weight, C-reactive protein [CRP], fecal calprotectin) were summarized descriptively.

Results

- In total, 1,176 patients in RCTs and 5,267 in RWS were treated with the bDMARD.
- On average, RWS patients were older, had longer disease duration, and had higher CRP, among other differences, compared with RCT patients (**Table 1**).

Table 1. Baseline characteristics

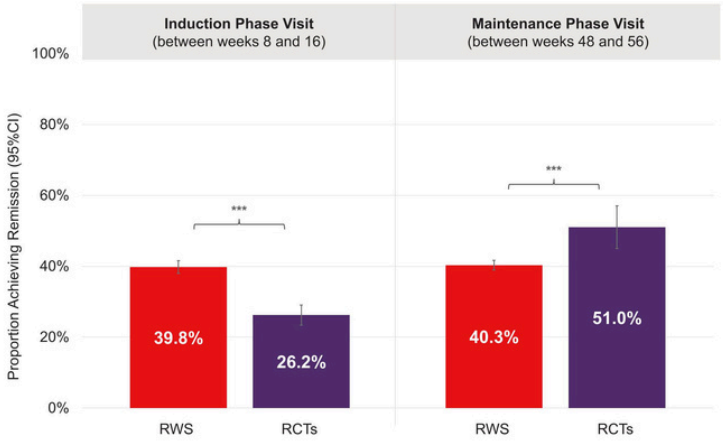
Parameter	19 Large-scale RWS		3 Pivotal Phase 3 RCTs		*P*-value	Percentage Difference, RWS vs RCT
	Proportion of Patients	n	Proportion of Patients	n		
Male sex (%)	45.8%	4,933	43.1%	1,176	0.0913	6.3%
Prior TNF (%)	84.2%	4,593	65.9%	1,176	**<0.001**	27.8%
Immunosuppressants (%)	32.6%	4,145	34.4%	1,176	0.2675	-5.0%
Corticosteroids (%)	39.3%	3,876	44.5%	1,176	**0.0015**	-11.7%
Montreal Classification						
L1 (%)	27.5%	4,339	18.9%	1,176	**<0.001**	45.8%
L2 (%)	17.6%	4,462	18.3%	1,176	0.5642	-3.9%
L3 (%)	52.1%	4,462	62.5%	1,176	**<0.001**	-16.6%
L4 (%)	10.8%	2,663	17.9%	1,176	**<0.001**	-39.8%
P (%)	31.6%	4,462	35.7%	1,176	**0.0076**	-11.5%
	Mean	n	Mean	n		Percent difference, RWS vs RCT
Age (years)	40.4	5,144	38.0	1,176		6.2%
Disease duration (years)	11.7	4,878	10.5	1,176		11.7%
Weight (kg)	69.4	672	70.7	1,176		-1.9%
CRP (mg/L)	9.6	1,582	9.0	1,176		6.9%
Fecal calprotectin (mg/kg)	495.6	1,263	504.8	1,176		-1.8%

Abbreviations: CRP = C-reactive protein; RCT = randomized controlled trial; RWS = real-world studies; TNF = tumor necrosis factor. *P*-values calculated using two-sample z-tests for proportions. Bolded *P*-values indicate statistical significance, defined as *P*<0.05. The Montreal Classification of CD defines disease location as L1 (ileum only), L2 (colon only), L3 (ileocolonic), and L4 (upper gastrointestinal/proximal involvement); perianal disease is denoted as "P". Immunosuppressants included azathioprine, mercaptopurine, and methotrexate.

Results (cont.)

- Clinical remission rates during induction were significantly higher in the RWS (39.8%) compared to the RCTs (26.2%; P<0.001). However, clinical remission rates during the maintenance phase were significantly lower in the RWS (40.3%) versus RCTs (51.0%, P<0.001) (**Figure 2**).

Figure 2. Remission outcomes



Abbreviations: RCT = randomized controlled trial; RWS = real-world studies
*** = *P*<0.001. *P*-values calculated using two-sample z-tests for proportions. Error bars represent 95% CI. Number of patients included in the analysis: from left to right: n=2,899; n=912; n=4,654; n=264.

Discussion

- This comparison of RCTs and RWS evaluating a bDMARD in CD revealed important differences in:
  - Baseline patient characteristics
  - Rates of achieving remission
- The differences observed likely reflect the broader inclusiveness of RWS.
- RCT data suggested continued improvements in remission achievement after induction (>16 weeks).
  - The results from the RWS suggested that approximately the same proportion of patients would achieve remission through induction as through the first year, underscoring the importance of early outcomes.
- Limitations of this analysis include:
  - Focus on one bDMARD, meaning generalizability to other bDMARDs is unknown
  - Reliance on aggregated study-level results (rather than patient-level data)
  - Variation among RWS, including the timing of follow-up visits and reporting of baseline characteristics

Conclusions

- **The significant differences in patient characteristics and clinical outcomes between RWS and RCTs highlight the necessity of real-world evidence in a comprehensive assessment of marketed and new treatments.**
- **RWS provide important insights into patient populations and clinical effectiveness that differ from those assessed by RCTs, offering a complementary role in understanding the true therapeutic potential and limitations of treatments in real-world practice.**

References

1. Ingrasciotta Y, et al. *Pharmacol Res*. 2024 Feb;200:107074.
2. Vieujean S. et al. *Lancet Healthy Longev*. 2022 May;3(5):e356-e366.

Disclosures

NRB, **MB**, and **JS** are employees of PPD™ Observational Studies, Thermo Fisher Scientific. Funding for this poster was provided by Thermo Fisher Scientific.

Acknowledgments

Editorial and graphic design support was provided by Caroline Cole and Shani Berger of Thermo Fisher Scientific.

Comparison of Randomized Controlled Trials (RCT) with Open-label, Single-cohort, Real-world Studies (RWS) in the evaluation of pharmacologic treatment of Non-small Cell Lung Cancer (NSCLC)

ABS-28

Richard LeBrun Trachy,[1] Vincent McCarty,[1] Neil R. Brett,[2] Marielle Bassel,[2] John Sampalis[1,2]

[1]McGill University, Montreal, Canada; [2]Thermo Fisher Scientific, Montreal, Canada

Background

- Randomized controlled trials (RCT) have been considered the gold-standard in evaluating treatment efficacy.
- However, in some cases of rare diseases and end-stage cancers, RCTs are not feasible or ethical.
- In recent years, real-world evidence has become increasingly important in the assessment of efficacy and safety of novel and existing treatments, addressing requirements for all stakeholders including regulators, patients, and healthcare providers.
- In many cases, the use of external control arms (ECAs) is an efficient way of conducting comparative analyses and complementing RCTs.
  - Regulatory agencies are increasingly more accepting of ECAs given certain requirements of quality, minimal bias, and comparability with the target or study population are demonstrated.
- There are concerns regarding the interpretation and validity of results from real-world, single-cohort studies as ECAs with respect to validity and representation of results that would be obtained in RCTs.

Objective

- **The aim of this study was to assess the validity of using ECAs derived from real-world studies in the assessment of efficacy of treatments for non-small cell lung cancer (NSCLC).**
  - **Efficacy outcomes assessed were objective response rate (ORR) and progression-free survival (PFS)**

Methods

- Literature search (PubMed)
  - ((((((Randomized controlled trial) OR RCT)) OR randomized)) AND ((NSCLC) OR non-small cell lung cancer).
  - Records were then excluded based on the criteria, yielding a final set of 385 RCTs.
- Single-cohort studies assessing any of the 21 treatment protocols used as experimental or control arms in the RCTs were identified.
  - ((((((Single-Arm) OR Single Arm)) OR nonrandomized)) AND ((NSCLC) OR non-small cell lung cancer). This search yielded 89 single-cohort studies.
- Finally, the comparable treatment were cross-verified to ensure it was evaluating the same treatment for single-cohort studies and RCTs.
- This resulted in 87 RCTs and 48 single-cohort studies used in the final evaluation for 21 treatments for NSCLC.

Results

Table 1. Summary of Patient Characteristics and Outcomes Across Treatments

Treatment Number	RCT					Single-cohort Study				
	N	Age (Years)[a]	Male %	ORR[a]	PFS[a]	N	Age (Years)[a]	Male %	ORR[a]	PFS[a]
1a	632	61	38	68	76	576	66	35	50	77
1b	32	67	84	9	8	18	39	33		11
2b	22	59	82	27		38	59	68	13	11
3a	3211	62	60	46	65	306	63	75	45	62
4a	75	67	40	69	90	109	66	39	77	84
4a	75	67	40	69	90	109	66	39	77	84
5b	39	61	50	64	39	76	63	59	20	33
6b	137	58	50	74	81	28	54	49	67	76
8a	60	58	93	42	54	10	77		50	20
9a	397	67	61	17	43	53	65	77	36	39
10b	115	54	41		42	227	53	47	55	65
12b	138	60	41	62	65	30	56	44	70	83
14b	699	64	65	20	35	68	61	57	31	38
15a	507	64	37	56	66	355	69	37	62	67
15b	1916	63	58	15	29	326	66	64		11
16b	434	66	57	29	44	43	57	30	5	26
17b	126	61	57	20	38	136	69	64	27	32
18b	765	61	71	19	33	53	61	62	9	21
19a	279	64	36	80	88	36	60	61	50	64
19b	384	60	40	68	78	228	65	30	69	70
20a	945	64	63	39	57	132	68	67	51	60
21a	168	71	60	22	36	66	76	65	14	35
21b	423	64	60	15	38	30	67	97		35

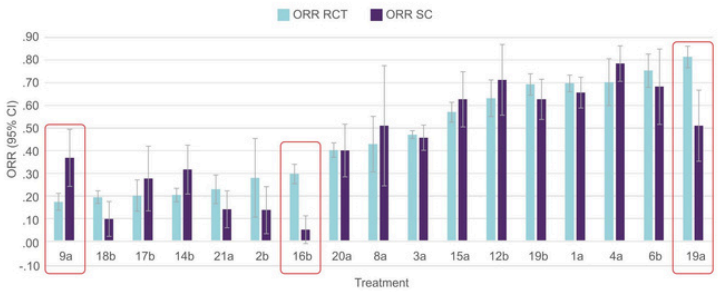
[a]Weighted Average
Abbreviations: ORR = objective response rate; PFS = progression-free survival; RCT = randomized controlled trial

Results (cont.)

Objective Response Rate (Figure 1)

- The ORR reported in RCTs and real-world studies was not statistically significant for 14/17 (82.4%) of treatments
- **The red boxes indicate ORR was significantly higher for the RCT in 2/17 (11.7%) and higher for real-world study in 1/17 (5.8%)**

Figure 1. ORR by Treatment and Study Type

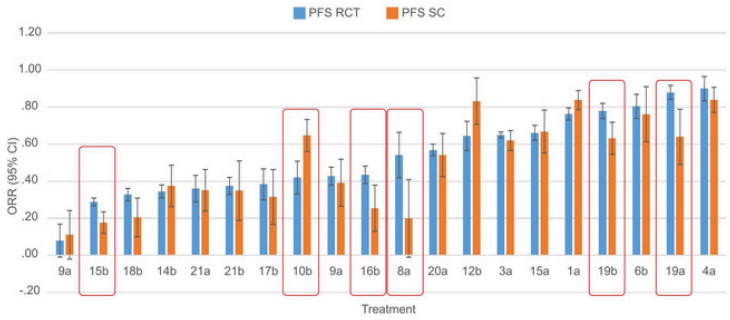


Abbreviations: ORR = objective response rate; RCT = randomized controlled trial; SC = single cohort; first-line treatment (a); subsequent-line treatment (b)

Progression-free Survival (Figure 2)

- PFS was not significantly different for 14/20 (70%) treatments.
- **The red boxes indicate PFS was significantly higher in the RCT for 6/20 (30%) treatments.**

Figure 2. PFS by Treatment and Study Type



Abbreviations: PFS = progression-free survival; RCT = randomized controlled trial; SC = single cohort; first-line treatment (a); subsequent-line treatment (b)

Discussion

- The results of this study demonstrated that the ORR and PFS observed in RCTs and single-cohort studies utilizing ECAs evaluating the treatment of NSCLC are comparable.

Limitations

- Focus on NSCLC: Generalization to other therapeutic areas requires further assessment.
- Publication Bias: Studies reported in the literature may be biased towards positive results.
- Design Limitations of single-cohort studies:
  - Random selection of patients could not be assessed
  - Timing of single-cohort study relative to the RCT was not assessed
  - Potential for bias by indication
  - Agreement with ECAs may be higher if single-cohort studies were specifically designed as ECAs

Conclusions

- **Well designed and carefully conducted single-cohort studies could provide ECAs replacing comparators in RCTs**
- **When fully randomized RCTs are not feasible or ethical, a real-world ECA would be a viable alternative.**
- **Real-world studies can enhance and supplement the evaluation of treatments for rare diseases or advanced cancers.**

Disclosures

NRB, **MB**, and **JS** are employees of PPD™ Observational Studies, Thermo Fisher Scientific. Funding for this poster was provided by Thermo Fisher Scientific.

Acknowledgements

Editorial and graphic design support were provided by Michael Grossi and Kawthar Nakayima of Thermo Fisher Scientific.

Presented at the GetReal Conference 2025 | 1–2 October 2025 | Utrecht, The Netherlands

Utility of Real-world External Control Arm (ECA) in the Evaluation of Pharmacologic Treatment of Non-small Cell Lung Cancer (NSCLC)

ABS-29

Richard LeBrun Trachy,[1] Vincent McCarty,[1] Neil R. Brett,[2] Marielle Bassel,[2] John Sampalis[1,2]
[1]McGill University, Montreal, Canada; [2]Thermo Fisher Scientific, Montreal, Canada

Background

- Randomized controlled trials (RCT) have been considered the gold standard in evaluating treatment efficacy.
- However, in some cases RCTs are not feasible or ethical, like with rare diseases and end-stage cancers.
- In recent years, real-world evidence has become increasingly important in the assessment of the efficacy and safety of novel and existing treatments, addressing requirements for all stakeholders, including regulators, patients, and healthcare providers.
- In many cases, the use of external control arms (ECAs) is an efficient way of conducting comparative analyses and complementing RCTs.
  - Regulatory agencies are increasingly more accepting of ECAs given certain requirements of quality, minimal bias, and comparability with the target or study population are demonstrated.
- There are concerns regarding the interpretation and validity of results from real-world, single-cohort (SC) studies with respect to validity and representation of results that would be obtained in RCTs.

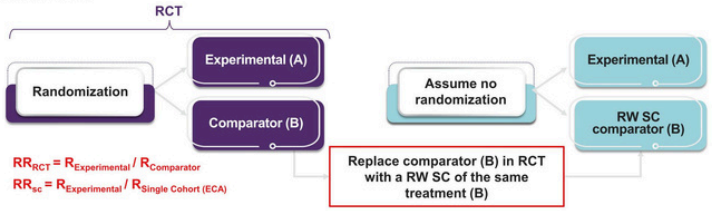
Objective

- **The objective of this study was to compare the results observed in RCTs and real-world studies (RWS) assessing the efficacy of treatments for non-small cell lung cancer (NSCLC).**
- **The efficacy outcomes assessed were objective response rate (ORR), progression-free survival (PFS) and overall survival (OS).**

Methods

- The following search string was used for the literature search in PubMed:
  - (((Randomized controlled trial) OR (RCT)) OR (randomized)) AND ((NSCLC) OR (non-small cell lung cancer).
- Records were then excluded based on the criteria, yielding a final set of 385 RCTs.
- Single cohort (SC) studies assessing any of the 21 treatment protocols used as experimental or control arms in the RCTs were identified using the following search string:
  - (((Single-Arm) OR (Single Arm)) OR (nonrandomized)) AND ((NSCLC) OR (non-small cell lung cancer)). This search yielded 89 SC studies.
- Finally, the comparable treatment was cross-verified to ensure the same treatment was evaluated for both SC and RCTs.
- This resulted in 87 RCTs and 48 SC studies used in the final evaluation of 21 treatments for NSCLC.

Figure 1. Design schematic for using RWS as comparators to RCT experimental treatments



Abbreviations: RCT = randomized controlled trial; RR = risk ratio; RW = real world; SC = single cohort
Treatments included: afatinib, apatinib, bevacizumab, brigatinib, carboplatin, ceritinib, cisplatin, crizotinib, dabrafenib, durvalumab, erlotinib, nab-paclitaxel, nivolumab, osimertinib, pembrolizumab, pemetrexed

Results

Table 1. Patient characteristics and outcomes

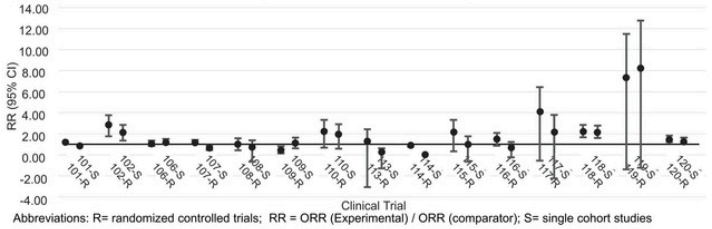
Treatment Number	RCT					SC Study				
	N	Age (Years)[a]	Male %	ORR[a]	PFS[a]	N	Age (Years)[a]	Male %	ORR[a]	PFS[a]
1a	632	61	38	68	76	576	66	35	50	77
1b	32	67	84	9	8	18	39	33		11
2b	22	59	82	27		38	59	68	13	11
3a	3211	62	60	46	65	306	63	75	45	62
4a	75	67	40	69	90	109	66	39	77	84
4a	75	67	40	69	90	109	66	39	77	84
5b	39	61	50	64	39	76	63	59	20	33
6b	137	58	50	74	81	28	54	49	67	76
8a	60	58	93	42	54	10	77		50	20
9a	397	67	61	17	43	53	65	77	36	39
10b	115	54	41		42	227	53	47	55	65
12b	138	60	41	62	65	30	56	44	70	83
14b	699	64	65	20	35	68	61	57	31	38
15a	507	64	37	56	66	355	69	37	62	67
15b	1916	63	58	15	29	326	66	64		11
16b	434	66	57	29	44	43	57	30	5	26
17b	126	61	57	20	38	136	69	64	27	32
18b	765	61	71	19	33	53	61	62	9	21
19a	279	64	36	80	88	36	60	61	50	64
19b	384	60	40	68	78	228	65	30	69	70
20a	945	64	63	39	57	132	68	67	51	60
21a	168	71	60	22	36	66	76	65	14	35
21b	423	64	60	15	38	30	67	97		35

[a]Weighted average
Abbreviations: ORR = objective response rate; PFS = progression-free survival; RCT = randomized controlled trial; SC = single cohort

Results (cont.)

- For ORR, there was agreement between the RCTs and ECA-RCTs for 9/15 (60%) in **Figure 2**.
- The purple box indicates non-significant different direction (NSDD) in 5/15 (33.3%).
- The red box indicates significant in different direction (SDD) in 1/15 (6.7%).

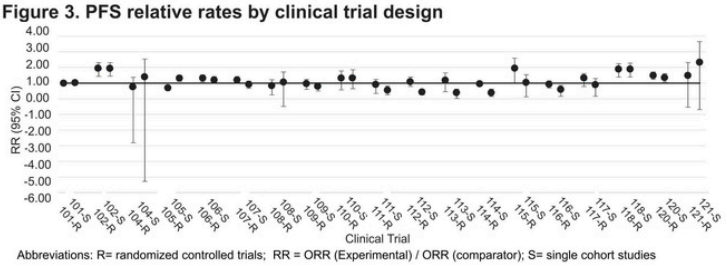
Figure 2. ORR relative rates by clinical trial and design



Abbreviations: R= randomized controlled trials; RR = ORR (Experimental) / ORR (comparator); S= single cohort studies

- For PFS, there was agreement between the RCTs and ECA-RCTs for 12/15 (80.0%) in **Figure 3**.
- The purple box indicates NSDD in 2/15 (13.3%).
- The red box indicates SDD in 1/15 (6.7%).

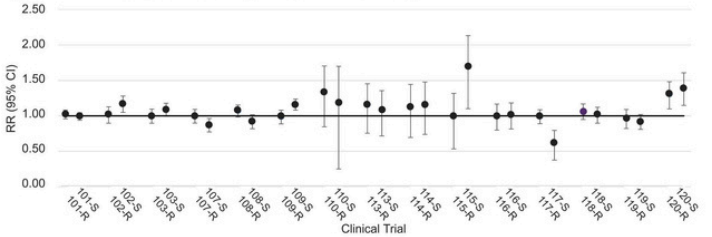
Figure 3. PFS relative rates by clinical trial design



Abbreviations: R= randomized controlled trials; RR = ORR (Experimental) / ORR (comparator); S= single cohort studies

- For OS, there was agreement (A) between the RCTs and ECA-RCTs for 14/15 (93.3%) in **Figure 4**.
- The red box indicates SDD in 1/15 (6.7%).

Figure 4. OS relative rates by clinical trial design



Abbreviations: R= randomized controlled trials; RR = ORR (Experimental) / ORR (comparator); S= single cohort studies

Discussion

- The results of this study demonstrated that the ORR, progression-free survival (PFS), and OS were comparable in RCTs and SC studies using ECAs and evaluating the treatment of NSCLC.

Limitations

- Focus on NSCLC: Generalization to other therapeutic areas requires further assessment.
- Publication bias: Studies reported in the literature may be biased towards positive results.
- Design limitations of SC studies:
  - Random selection of patients could not be assessed.
  - The timing of SC studies relative to the RCTs was not assessed.
  - There is potential for bias by indication.

Conclusions

- **Well-designed and carefully conducted SC studies can supplement the evidence provided by RCTs.**
- **When RCTs are not feasible or ethical, the SC design can be a viable alternative.**
- **RWS can enhance and supplement the evaluation of treatments for rare diseases or advanced cancers.**

Disclosures

NRB, **MB**, and **JS** are employees of PPD™ Observational Studies, Thermo Fisher Scientific. Funding for this poster was provided by Thermo Fisher Scientific.

Acknowledgements

Editorial and graphic design support was provided by Caroline Cole and Karissa Calara of Thermo Fisher Scientific.

Advancing Real-World Evidence with Causal AI

A SCALABLE FRAMEWORK FOR UNLOCKING ACTIONABLE INSIGHTS FROM REAL-WORLD PATIENT DATA

Andreas Nierlich, andreas.nierlich@xplain-data.com, Xplain Data GmbH, Zorneding, Germany

Standard Adjustment vs. Causal Discovery – What´s the Difference?

Standard Adjustment / Conventional Methods

Complex Data > "Flattened Tables"

Require assumptions & predefined hypotheses

vs

Causal Discovery on ObjectAnalytics

Complex Data > Preserved Richness

Unbiased search without hypotheses

Conventional methods for estimating intervention effects (e.g., propensity scores, regression, stratification) depend on strong **assumptions**:

- They require **predefined hypotheses** about interventions (you cannot simply ask "which factors matter?").
- **Confounders** must be explicitly specified for adjustment, introducing further assumptions.

Assembling corresponding variables in a flat table causes the loss of rich information from the original, complex patient data.

Causal Discovery – A Different Approach:

- **Works directly with complex real-world data** (e.g., EMRs with hundreds of interrelated tables), preserving the full richness of information.
- Enables **open-ended exploration:** "Show me all causal pathways leading to the outcome of interest."

Causal Discovery represents an unbiased search for causal factors and confounders within the full complexity of real-world data.

Methodology: ObjectAnalytics – Enabling Deep Causal Search

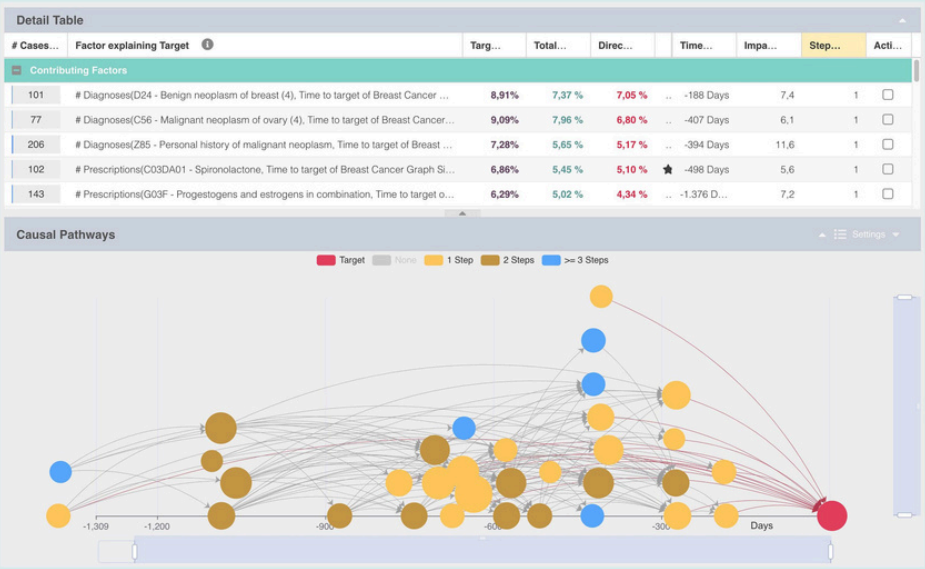


Figure 1. Example Results: Causal network for target breast cancer (DAG with direct, indirect, and total effects).

Comprehensive patient information is necessary for identifying **all** factors and confounders. This information can span hundreds of interrelated tables.

Challenge:
Causal algorithms must operate in this complex environment.

Solution: ObjectAnalytics –
transforms fragmented RWD into holistic patient objects:

- **Patented, object-centered model** integrates fragmented, longitudinal data from multiple tables and sources (RWD) into a unified 360° patient view.
- All patient-related information is **linked in an object tree**: patient as root, with diagnoses, treatments, prescriptions, etc. as sub-objects.
- **Overcomes the flat-table limitations** of relational databases.
- Provides the **foundation for Causal Discovery** algorithms to conduct deep, automated confounder search **without prior assumptions.**

Examples of Causal Discovery Application in Healthcare

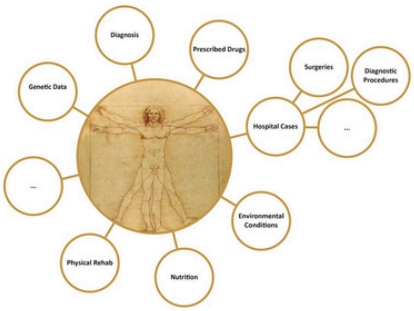


Figure 2. Object-centric patient view: foundation for deep causal search by rapidly generating and testing millions of hypotheses for factors and confounders.

Market Access: Causal Drivers in the Therapy Journey

- **Therapy starts & restarts:** Analyze the causal drivers behind *why* patients choose Product X over competing products.
- **Switch Wins | Losses:** Comprehend market dynamics and reasons for switching.

Real World Evidence: Causal Pathways in the Diseases Journey

- **Analyze typical disease pathways:** prior conditions, progression, escalation.
- **Understand causal factors:** triggers of disease onset or relapse
- **Perceive entire causal pathways:** direct & indirect drivers across multiple hubs.
- **Perform deep confounder search:** robust drug-effect evaluation in RWD studies.

Treatment Effectiveness & Optimization:

- **Predict response vs. non-response** based on patient characteristics.
- Determine causal factors of **dose optimization** and **titration success**.
- Analyze causal interactions and synergies in **combination therapies.**

Other applications:

Healthcare resource utilization, safety and pharmacovigilance, precision medicine, and biomarker discovery, among others.

RELEVANCE FOR LIFE SCIENCE ORGANIZATIONS AND RESEARCHERS

- **Faster evidence generation:** uncover causal drivers without months of manual hypothesis testing.
- **More robust RWE studies:** unbiased adjustment for confounders improves credibility with regulators and payers.
- **Deeper market insights:** identify true drivers of therapy choice, switching, and adherence.
- **Improved treatment optimization:** personalize dosage, predict response, and optimize drug combinations.
- **Competitive advantage:** leverage the full richness of RWD while competitors rely on limited regression-based methods.
- **Scalability:** a framework that can handle millions of patient records and hundreds of variables simultaneously.

Ask for a live demo.

Video Intro:
Causal Discovery on Patient Data

Xplain Data · info@xplain-data.com · www.xplain-data.com