Medical Writing

ALSO IN THIS ISSUE...

- Remembering Keith Veitch
- AI-enabled medical writing
- The microbiota connection

Volume 34 Number 3 | **September 2025**

EMWA EUROPEAN MEDICAL WRITERS ASSOCIATION

Medical Writing is the official journal of the European Medical Writers Association (EMWA). It is a quarterly journal that publishes articles on topics relevant to professional medical writers. Members of EMWA receive *Medical Writing* as part of their membership. For more information, contact mew@emwa.org.

Medical Writing

THIS ISSUE **September 2025** | Volume 34 Number 3

Real World Data/Real World Evidence

"Real-world data (RWD) and real-world evidence (RWE) are no longer auxiliary concepts – they are becoming central to how we understand health, disease, and care."

Medical Writing Guest Editors
Laura C. Collada Ali and Maria Kołtowska-Häggström

Real-world data and real-world evidence: A new ethics of knowing

doi: 10.56012/rfcq2943

The field of medicine stands at a critical juncture. In the ever-evolving journey from evidence to action, from clinic to community, and from regulation to lived experience, real-world data (RWD) and real-world evidence (RWE) are no longer auxiliary concepts – they are becoming central to how we understand health, disease, and care. This special issue of EMWA's *Medical Writing* journal is dedicated to exploring the technical and practical dimensions of RWD/RWE, recognising it as a powerful conduit for patient-centered insight and innovation grounded in reality, responsive to needs, and driven by practicality.

We are no longer satisfied with understanding health solely through the lens of randomised controlled trials (RCTs), valuable though they are. While RCTs answer the "does it work" question, RWD/RWE tell us how it works in real-life. Thus, now we now ask ourselves: How do patients experience treatment in the everyday world? What happens between visits, outside hospitals, beyond clinical endpoints? RWD/RWE shifts our focus from the controlled to the lived, from the prescriptive to the observed, and perhaps most importantly, from the theoretical to the ethical. It draws us into the reality of care – messy, complex, and often unmeasured – and forces us to confront whether our systems reflect the needs of the people they serve.

Image: Freepik

GUEST EDITORS

Laura C. Collada Ali
laura.collada@trilogywriting.com

Maria Kołtowska-Häggström
maria.koltowska-haggstrom@propermedicalwriting.com

Ethics: Who owns the evidence?

At the heart of RWD/RWE lies a fundamental ethical question: Whose knowledge counts? When we speak of patient-generated data – whether from electronic health records, mobile health devices, or social determinants of health – we are not merely gathering statistics. We are witnessing people's lives in data form. And with that comes the imperative to respect privacy, autonomy, and agency.

RWE gives rise to a new ethics of knowing, which rethinks the principles, responsibilities and power structures involved in the creation and dissemination of knowledge. It challenges us to recognise that evidence is not neutral – it is shaped by who collects it, how it is interpreted, and for what purpose. As one of our feature articles explores from a European regulatory perspective, the development of guidance around RWD/RWE is not just about technical rigour; it's about safeguarding trust. Transparency, inclusivity, and participatory design must become ethical cornerstones of how we integrate RWE into health systems – not only to reflect scientific rigor, but also the realities of care delivery and decision-making.

Patient-centeredness: Evidence that feels real

Patient-centered care is often defined in philosophical terms. RWD/RWE makes it tangible. We see this in our contributors' work on chronic disease management, digital therapeutics, and community health analytics. By incorporating diverse data sources – wearables, patient-reported outcomes, behavioural data – we begin to understand not just how treatments work, but how they work for real people in real contexts.

What emerges is a more holistic picture of health. Lifestyle factors, social environments, and personal goals – all too often left out of traditional clinical data – are captured in ways that illuminate the true burden and benefits of care. Several articles in this issue address the intersection of RWD with lifestyle medicine, showing how real-world insights can lead to proactive, preventive strategies that align with patients' needs – and not just clinical endpoints.

Crucially, RWD/RWE allows us to hear from populations often excluded from clinical trials: older adults, people with comorbidities, rural communities, and ethnically diverse groups. Patient-centeredness means acknowledging these voices not as exceptions, but as essential to understanding effectiveness in the real world. The data must reflect the lived realities of those whose health and outcomes depend on our systems.

Opportunity: From observation to transformation

The promise of RWD/RWE is immense — but so is the responsibility. We have the tools to detect patterns across vast datasets, to tailor interventions to specific populations, and to predict outcomes with increasing precision. But will we use this opportunity to reinforce equity, or will we widen gaps?

EUROPEAN MEDICAL WRITERS ASSOCIATION

Submissions:
For instructions to authors, go to the journal section of EMWA's website (www.journal.emwa.org).
All manuscripts should be submitted to mew@emwa.org.

Amid all this, medical communicators have a vital role to play. We are not merely translators of data – we are stewards of meaning.

One of the most compelling aspects of this issue is how it spans the full ecosystem – from the regulator's office to the patient's kitchen table. We see articles addressing how the industry is embracing RWD and its potential, the future of drug safety, statistical and reporting challenges, AI-driven insights, medical device surveillance, language and evidence in a polarised era, and the practical relevance of lifestyle medicine. Together, they illustrate that RWD/RWE is not just a technical evolution; it's a cultural shift in how we think about evidence – one that emphasises practicality, responsiveness, and accountability.

In health economics, RWE is helping systems allocate resources based on actual utilisation patterns. In rare diseases, RWD offers the only practical route to understanding natural history and treatment response. In pharmacovigilance, RWE captures safety signals that would be missed in trials. In each case, the opportunity is not just to observe, but to act – to refine policy, improve practice, and empower patients.

The role of the medical writer

Amid all this, medical communicators have a vital role to play. We are not merely translators of data – we are stewards of meaning. In the RWD/RWE landscape, clarity, relevance, and practical insight matter more than ever. It is up to us to help stakeholders – clinicians, patients, payers, regulators – understand what this data means and what it does not. We must advocate for ethical reporting, resist oversimplification, and emphasise the needs and narratives behind the statistics.

Moreover, we must be champions of inclusivity in how evidence is communicated. As more patient-generated data enters the evidence base, we must ensure that patients are not just sources of information, but co-authors of understanding. Their lived realities must shape not just the data we collect, but the decisions we make and the words we choose.

Looking ahead

This issue reflects a collective recognition: that evidence must meet life where it happens. As we move forward, let us centre our work on a simple, radical idea – that real-world evidence is not just about what works, but about what matters. It must address real needs, reflect lived realities, and be applied with practical wisdom.

To all our readers – medical writers and science communicators in Europe and beyond – we invite you to consider how RWD/RWE can be a tool not only for insight but for integrity. Let us use this moment to reimagine evidence as not just a pathway to approval, but a commitment to accountability and care.

Let the real world speak – and let us listen, wisely and well.

Author information

Laura C. Collada Ali is a Senior Medical Writing Manager at Trilogy Writing & Consulting. She has over 25 years of experience writing, editing, and managing medical and scientific documentation for diverse clients and audiences worldwide. She is a member of EMWA's Executive Committee and a workshop leader and holds multiple certificates in linguistics, pharmacology, medical devices, clinical research, and regulatory writing.

Maria Kołtowska-Häggström, MD, PhD, runs Proper Medical Writing, the first Polish medical writing agency. She has 30 years of experience in the pharmaceutical industry, primarily in clinical research, including real-world data (RWD) and real-world evidence (RWE). Maria has authored over 80 peer-reviewed publications, many focused on RWE and patient-reported outcomes. She is a member of EMWA, EASE, ESE, and GRS; serves as a journal reviewer and Associate Editor for BMC Endocrine Disorders; and is an EMWA workshop leader, and a section editor for *Medical Writing*. She was EMWA president in 2023-2024.

President's Message

Taking our cues from the real world

Martin Delahunty
EMWA President 2025–26
president@emwa.org

doi: 10.56012/jruo5220

Workshop leaders at EMWA's Spring 2025 conference in Riga, Latvia

Dear EMWA Members and Colleagues,

Welcome to this September 2025 issue of *Medical Writing*, focusing on the increasingly important topic of real-world data and real-world evidence. I want to extend my sincere gratitude to our guest editors, Maria Kołtowska-Häggström and Laura C. Collada Ali, for bringing together such a comprehensive collection of articles on this critical subject.

I also want to thank our *Medical Writing* Editor-in-Chief, Raquel Billiones, and the Editorial Board – all EMWA volunteers working tirelessly to commission hugely valuable and practical content for our members.

Before exploring this issue's theme, I must acknowledge the profound loss our community experienced this summer with the passing of Keith Veitch on July 29, 2025. Keith served as EMWA President from 2000 to 2001 and was a cherished member of our community for decades. His contributions to medical writing, leadership, and infectious spirit touched countless lives within EMWA and beyond. I encourage you to read the full memorial article in this issue, which beautifully captures Keith's professional legacy and the warmth he brought to our community.

Real-world data and real-world evidence represent one of the most significant developments in modern medical research and healthcare decision-making. As medical writers, we find ourselves at the forefront of communicating insights derived from real-world settings, where the complexities of everyday clinical practice often reveal different patterns than those observed in controlled trial environments. This evolution demands not only technical expertise but also a nuanced understanding of how to present real-world findings in ways that are both scientifically rigorous and accessible to diverse audiences.

Our milestone 60th EMWA Conference, to be held virtually November 14–29, will feature an impressive programme with 35 workshops, 5 seminars, our Freelance Business Forum, and exclusive sponsored sessions. The virtual format allows us to reach members worldwide and provide the flexibility many of you need to participate fully in this special anniversary conference.

Integral to the success of our conference education programme are our workshop leaders who share their knowledge and expertise to advance the profession of medical writing.

We're already looking forward to the 61st EMWA Spring Conference, which will take place in an in-person format from May 5–8, 2026, in the wonderful city of Barcelona. Please mark your diaries now, and watch for further announcements toward the end of this year. There's something truly special about gathering in person, and Barcelona promises to provide an inspiring backdrop for learning, networking, and celebrating our community. We will also be celebrating the 10th anniversary of CORE Reference.

As we continue to navigate the evolving landscape of medical writing – from real-world evidence to new regulatory requirements and emerging technologies – I'm continually impressed by the adaptability and expertise of our EMWA community. Your commitment to excellence in medical communication makes a real difference in advancing healthcare and improving patient outcomes.

Thank you for your continued engagement with EMWA and with *Medical Writing*. I hope you find this issue both informative and practical as you tackle real-world data challenges in your own work.

Best regards.

Editorial Board meeting in Riga

From The Editor

Raquel Billiones
Editor-in-Chief
editor@emwa.org
0000-0003-1975-8762

doi: 10.56012/ektm6292

The Editor's Red Pencil: A sad goodbye to Keith Veitch, and *Medical Writing* goes fully digital

This is our first fully digital, paperless issue of *Medical Writing*. It is tragically unexpected that with this milestone issue, in which we look back at our origins, we also say goodbye to our dear colleague Keith Veitch, who died on July 29. (Please see his memorial on p. 8).

You see, not only was Keith a former EMWA President, he was one of the pioneer editors of the EMWA publication, then called the *EMWA Newsletter*, which he edited from 1996 to 1998.

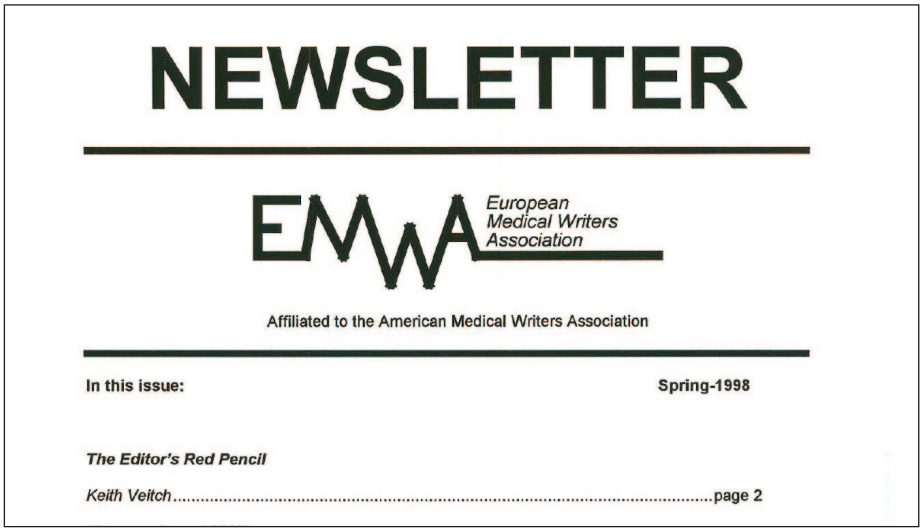
When I joined EMWA in 2006, the name Keith Veitch was already a legend. But funnily enough, despite our common EMWA roots, I met Keith, not through or within the organisation, but when I transitioned from the CRO world to big pharma in 2019. I joined Takeda Vaccines as regulatory writing head, and he was there, the publications writing head, ready to welcome me. Full of warmth and a wacky sense of humour, I was so glad to have a fellow medical writer help me navigate the corporate world. I especially remember his stories about his daughters (he was such a doting father) and the early days of the EMWA (there's a connection, you'll see).

When he took on the task of being newsletter editor, he wrote *"… this should be an easy job because all those professional writers out there will be able to provide tons of well-written, easily edited material for inclusion. Go ahead, prove me right!"*[1]

But he was proven wrong:

"It has always amazed me that an organisation composed of people who make their living from writing, often unacknowledged, are not eager to apply their skills to contribute to their own newsletter… "[2]

And like many of us, EMWA sometimes encroached into our private lives and Keith's daughters supported the organisation without even realising it. *"… My kids were asking when they have to fold papers and stuff envelopes again,"* he wrote.[3]



NEWSLETTER

EMWA European Medical Writers Association

Affiliated to the American Medical Writers Association

In this issue: Spring-1998

The Editor's Red Pencil

Keith Veitch.....page 2

Little did I know that I would follow in his footsteps and become editor in 2021.

Well, Keith would be happy to see that during the last few years we have (more than) enough content to fill each issue. And starting this autumn of 2025, we are completely paper- and envelope-free.

Keith, we dedicate this issue to you as I use your red pencil one last time.

What is your role in the data universe?

This issue on real-world data/real-world evidence (RWD/RWE) reminds me of the June 2020 issue on The Data Economy which I guest edited, together with Sam Hamilton. In our editorial, we wrote "*data are economic assets that power the so-called fourth industrial revolution*" where medical writers and communicators have a vital role to play.[4] And what would that role be?

As an academic, I was a data collector, be it collecting field samples or conducting experiments in the lab. As a medical writer and clinical researcher, I have always been a data processor of patient health data in the documents I author.

Recently, I participated in the SwissHeart study – as a data subject.

"The SwissHeart Deep Imaging Data Collection (DIDC) Study records MRI data of the heart as well as information on lifestyle, blood values, body composition and fitness of a representative population living in Switzerland. Based on the information collected, digital models of the heart and vessels will be created and made available in an anonymised form to help improving the understanding of the development, diagnosis and prevention of cardiovascular diseases."[5]

Going through the informed consent process, and then on to answering questions on my medical history, and then undergoing the

assessments, I got to experience how to be a study participant. I marvelled at my sense of fulfilment when I met the eligibility criteria to contribute my very own health data to this body of RWD of the Swiss population. Plus, the medical device geek in me really enjoyed the nifty gadgets used in the study.

Yet, as with all precious commodities, data also have a dark side, especially in this age of AI. Everyday, we hear reports of data being faked, manipulated, misused, stolen, deleted. So let us be vigilant and remember another very important role for us medical writers and communicators – as data stewards to help protect quality and integrity of the data we are dealing with.

To close, here's one last quote from Keith:

"This Newsletter is for you, about you, and written by you … So let's make this a Newsletter to be proud of."[1]

As *MEW* enters its fully digital era, kudos to all those who contributed to this fantastic issue on RWD/RWE, especially our guest editors Laura Collada Ali and Maria Kołtowska-Häggström and all those working behind the scenes to make *MEW* a publication we all can be proud of.

References

1. Veitch K. The Editor's Red Pencil. EMWA Newsletter. Summer/Autumn 1996.
2. Veitch K. The Editor's Red Pencil. EMWA Newsletter. Spring 1998.
3. Veitch K. The Editor's Red Pencil. EMWA Newsletter. Summer/Autumn 1997.
4. Billiones R, Hamilton S. Editorial, The data economy. Available from: https://journal.emwa.org/the-data-economy/
5. University of Zurich. SwissHeart Study. Available from: https://cmr.ethz.ch/swiss-heart-study.html

Introducing the MEW Flipbook

exclusively for EMWA members

From Sep 2025, MEW is available as a flipbook

A flipbook is an interactive digital publication with a realistic page-flip effect that makes it look just like a printed copy. A 'flipbook PDF' is not a distinct file format, but rather a static PDF document that has been converted into an interactive digital flipbook using specialised software or online tools. The digital flipbook uses animation to simulate the tactile experience of a physical book, with pages that appear to turn, complete with shadows and sound effects, offering a more engaging and interactive experience than a standard PDF.

Source: ChatGPT

In Memory of

Keith Veitch

Former EMWA President and cherished community member

It is with great sadness that EMWA announces the passing of Keith Veitch on July 29, 2025. Keith was not only a distinguished professional in medical writing and publications but also a cherished friend and mentor to countless members of the EMWA community.

Professional legacy

Keith's career journey exemplified the evolution of modern medical communications. Beginning as a biochemist, he successfully transitioned into publication writing over 30 years ago, starting with GSK in Belgium. His expertise and leadership quickly became evident as he progressed to head publication groups at Sanofi Pasteur in Lyon, France, and later at Novartis Vaccines in Amsterdam, the Netherlands.

Following his corporate career, Keith established a thriving freelance consultancy specialising in publications and medical writing, with expertise in infectious diseases and vaccines. His work contributed significantly to advancing standards in medical communication and ethical reporting practices.

EMWA leadership and contributions

Keith's dedication to EMWA was profound and lasting. Serving as President from 2000 to 2001, he helped shape the organization during a pivotal period in its development. As Editor of The EMWA Newsletter – one of the forerunners to *Medical Writing* – he contributed to establishing the foundation of EMWA's educational and communication efforts. (See p. 6)

Photo courtesy of Susan Bhatti

A personal memory – Susan Bhatti

I met Keith for the first time many years ago at an EMWA conference in Dublin. He was at the bar (of course) drinking Guinness (of course) and we chatted. It would be romantic to say it was love at first sight, but in fact our friendship and affection for each other grew during many EMWA conferences and over many years.

We discovered we had a lot in common and had shared memories of growing up in the UK, despite both of us living for decades in Europe (but in different countries). After a long-distance relationship for over 10 years, we were finally able to live together when Brexit forced the European Medicines Agency to move to Amsterdam, creating an opportunity for a role in my company in the country Keith had adopted as his home.

Keith filled my life with laughter and love. I miss him too much to explain. I am so grateful that EMWA enabled us to meet each other, and I am so blessed that he was able to share some of his life with me.

One of Keith's proudest moments was conferring the first EMWA Nick Thompson Fellowship, recognising his commitment to fostering excellence and supporting emerging talent in the field. His involvement extended beyond EMWA to organisations like The International Society for Publication Professionals (ISMPP) and The International Publication Planning Meeting (TIPPA), where he worked tirelessly to promote ethical reporting of clinical trials. As a member of the Good Publication Practices (GPP3) Steering Committee and contributing author, Keith helped establish guidelines that continue to influence medical writing practices today.

A spirit of joy and camaraderie

Those who knew Keith will remember him as much for his infectious personality as for his professional achievements. He brought warmth, humour, and an irreverent spirit to every EMWA conference and ISMPP gathering. Whether engaging in spirited late-evening discussions over beers or sharing his trademark pirate humour with close friends, Keith had an extraordinary ability to make colleagues feel welcomed and valued.

His openness in conversation and genuine interest in others created lasting friendships across the global medical writing community. Keith understood that professional excellence flourished best in an environment of mutual support and genuine camaraderie – values he embodied throughout his involvement with EMWA.

Lasting impact

Keith Veitch's legacy extends far beyond his impressive professional accomplishments. He leaves behind a community enriched by his mentorship, leadership, and friendship. His contributions to ethical medical writing standards, his dedication to EMWA's mission, and his ability to bring joy and authenticity to our professional gatherings will be remembered and cherished.

To his partner Susan, three daughters, three granddaughters and family and many friends around the world, we extend our deepest condolences. Keith's memory will live on in the standards he helped establish, the colleagues he mentored, and the countless moments of connection and laughter he shared with our community.

The medical writing world has lost a true champion, and EMWA has lost a dear friend whose spirit and contributions will be remembered for generations to come.

A personal memory – Art Gertel

I am very saddened, and shocked, by this news.

Keith was a good friend over the many years we shared EMWA membership and, as EMWA President, he was the one who conferred the first EMWA Nick Thompson Fellowship on Geoff Hall and me. He was irreverent and incorrigible, which made him all the more endearing. One year, he and I embarked on a road trip in the Big Sur area of California, and his thunderous snoring was the subject of much teasing (shared with Susan Bhatti, with whom he shared the latter part of his life). Keith and I also engaged in inside pirate joke humour, and often lapsed into Treasure Island vernacular, much to the embarrassment of our more professional colleagues.

I remember the ill-fated EMWA dinner boat cruise in the Malta harbour. The water was so rough that many guests requested to be offloaded and ferried back to shore. Of course, this maritime adventure triggered one-legged pirate "humour" between the two of us.

Keith will be missed, and the world is a darker place in his absence.

A personal memory – Julia Forjanic Klapproth

I am writing this with a heavy heart and deep gratitude to honour the life of a wonderful man – Keith Veitch. Keith was a man of wit – sharp, clever, and always ready with a turn of phrase that could make you laugh, think, or both at once. He had a gift for language, a true Geordie at heart, delighting in the playfulness of words and the power they held. Conversations with Keith were never ordinary; they sparkled with insight, humor, and warmth.

But Keith was more than just witty. He was thoughtful – always considering others, always listening with care. He was generous with his time, his wisdom, and his encouragement. And he was kind in the truest sense of the word: gentle, respectful, and deeply humane.

I had the privilege of serving as Vice President of EMWA when Keith was our President. He led with grace and clarity, never seeking the spotlight but always guiding with a steady hand. He believed in people. He believed in me. And that belief made all the difference.

Photo courtesy of Julia Forjanic Klapproth

This picture of the two of us was taken during the EMWA conference in Malta in 2005. Keith and I always had so much fun together and I love this picture because it captures something essential about Keith: his openness, his approachability, and the quiet joy he found in being part of a shared purpose. That image, like so many memories of him, is a snapshot of connection, of friendship, of trust.

Keith's legacy is not just in the roles he held or the words he wrote – it's in the lives he touched. In the laughter he sparked. In the confidence he nurtured in others. In the quiet, steadfast way he made things happen. We will miss him dearly.

EMWA News

SECTION EDITOR

Somsuvro Basu
basu.somsuvro@gmail.com

doi: 10.56012/frnz8066

EMWA Calendar

Check out the EMWA calendar, register, and schedule your favourite upcoming events: https://emwa.org/communities-engagement/emwa-events-calendar/

November 14–29, 2025

60th EMWA Conference.
Registration now open:
https://emwa.org/conference/

May 5–8, 2026

Spring Conference in Barcelona.
We have listened to your feedback and are adding some exciting and thought-provoking new sessions and activities.

EMWA on LinkedIn

Join our LinkedIn community and keep up to date:
https://www.linkedin.com/company/european-medical-writers-association-emwa-/posts/?feedView=all

Test your writing skills with the **EMWA Medical Writing Journal series**. Follow the **Behind the Conference series** of interviews with the EMWA Executive Committee.

Did you know?

Existing EMWA members can receive a 10% discount off their next year's subscription for referring a new member to EMWA.
For more information, please contact Head Office at info@emwa.org

Check out the back issues of EMWA's journal *Medical Writing* at https://journal.emwa.org!

CORE Reference

The CORE Reference Project is moving.

It is moving away from email and onto LinkedIn to streamline our distribution of educational materials, including the monthly News Summaries: https://www.linkedin.com/company/the-core-reference-project/

News Summaries and useful information up to the end of 2024 are archived at: https://www.core-reference.org/news-summaries/

EMWA Professional Development Committee Webinar

EMWA webinars help members to develop skills and keep up to date with new or rapidly developing areas.

Most of our webinars are live, online seminars with the opportunity for participant interaction. Webinar access is reserved for EMWA members only and requires registration.

For the planned or past webinars, please refer to this page: https://emwa.org/education/emwa-webinars-programme/

EMWA Special Interest Groups

EMWA membership allows you to participate in any **Special Interest Group** (SIG) Meet and Share, even if you are not an active member of that SIG.

These events are announced in the EMWA newsletter and in a separate mailing closer to the event date. The Meet and Share sessions are great opportunities to learn more about a particular topic in an informal setting. Some sessions may be recorded, but many are not.

SIG members, on the other hand, participate in all SIG meetings (as their availability permits) and/or are more involved in the SIG activities, requiring an active role in providing more in-depth knowledge about what is going on in the SIG area.

If you are interested in knowing more about the SIGs, please read this: https://emwa.org/communities-engagement/find-communities/special-interest-groups-working-groups/

EMWA Sponsorship

Is your company interested in becoming an EMWA sponsor?

Sponsoring EMWA offers valuable benefits, including direct access to key opinion leaders in medical writing, visibility at our conferences – the largest gatherings of medical writers in Europe – and ongoing professional exposure.

Our tailored partnership models are designed to suit a variety of promotional budgets. Get in touch today to explore how your company can benefit: emwa.org and info@emwa.org

Real-world evidence and real-world data in contemporary healthcare: A critical overview

Andrew Balkin
Ambermed Medical Writing and Consulting
London, United Kingdom

doi: 10.56012/cyuu7806

Correspondence to:
Andrew Balkin
andrew@ambermed.org

Abstract

The utilisation of real-world data (RWD) and real-world evidence (RWE) is transforming healthcare research and decision-making. As an alternative to traditional clinical trials, RWD encompasses data from everyday clinical settings, while RWE represents the actionable insights generated from analysing these data. The integration of RWD and RWE offers critical benefits for healthcare delivery, clinical decision-making, pharmaceutical development, and policy formulation. This article delves into the definitions of RWD and RWE, outlines their distinctions, explores their sources, and discusses their broad applications in modern healthcare. In doing so, it highlights the evolving role of RWD in shaping personalised, evidence-based care and the future of healthcare systems globally.

Introduction

In the last few decades, healthcare research has witnessed a transformative shift. While controlled clinical trials have historically been the gold standard for evaluating medical interventions, they are limited in their ability to reflect the complexity of real-world patient care. Real-world data (RWD) and real-world evidence (RWE) offer a solution by providing insights into healthcare practices, patient outcomes, and treatment effectiveness outside the highly controlled environments of clinical trials.[1] These data elements, when harnessed effectively, provide a more comprehensive understanding of healthcare delivery, driving better clinical decisions, more accurate policymaking, and more effective pharmaceutical development.[2]

RWD is the raw data collected from diverse sources, such as electronic health records (EHRs), insurance claims, patient registries, and wearable devices. RWE, on the other hand, refers to the actionable insights derived from systematically analysing RWD. Together, these tools are reshaping how the healthcare system makes decisions, implements treatments, and designs policies. This article explores the distinction between RWD and RWE, their sources, transformation processes, and their applications in improving healthcare outcomes.

Understanding real-world data (RWD)

RWD is defined as data collected from a variety of healthcare environments outside the scope of clinical trials. Unlike the carefully controlled settings of randomised controlled trials (RCTs), RWD captures the complexity of patient experiences in everyday clinical practice. It includes data from routine clinical care, patient registries, electronic health records, insurance claims, mobile health apps, wearable devices, and more. This data is valuable because it reflects how medical treatments, interventions, and healthcare systems work in real-world settings, as opposed to the idealised environments of clinical trials.

Key sources of RWD

RWD is collected from several diverse and interconnected sources (see Table 1). These data sources offer a comprehensive view of patient care and provide a wealth of information that can be used to evaluate healthcare trends and outcomes.[3]

Real-world evidence (RWE): From data to insight

RWE is the actionable insight that is derived from the analysis of RWD. While RWD represents the raw data collected from various sources, RWE is the end result of analysing this data in order to generate evidence that can inform clinical decision-making, healthcare policies, and pharmaceutical development (see Table 2).[4] The process of transforming RWD into RWE involves sophisticated methodologies that account for the inherent complexities and variability of RWD.

The process of converting RWD into RWE involves several stages. Each stage plays a crucial role in ensuring that the data is not only reliable but also actionable for clinical and policy decisions.[6] These stages are:

1. **Data collection**: The first step in generating RWE is the collection of data from multiple sources. This data includes not only clinical records but also information related to patient behaviour, socio-economic status, treatment adherence, and more.[1] The broad scope of data collected in real-world settings ensures that RWE captures the diversity and complexity of patient experiences.
2. **Data integration**: After data collection, the next step is to integrate and harmonise the various datasets. RWD often comes from disparate sources, and these sources may use different formats and standards.[3] The integration process ensures that all data are standardised and compatible, creating a unified dataset that can be analysed effectively.
3. **Data analysis and interpretation**: Once the data is integrated, it is analysed using advanced statistical techniques. Researchers apply machine learning models, regression analysis, and other methodologies to uncover patterns, trends, and relationships within the data.[6] The goal of this analysis is to extract meaningful insights that can guide clinical decisions, shape healthcare policies, and influence drug development strategies.
4. **Application of results**: The insights generated through RWE analysis are then applied in real-world healthcare settings. These insights can inform individual clinical decisions, enhance the design of healthcare policies, and improve the safety and efficacy of treatments.[5] By applying RWE, healthcare providers can better tailor treatment plans to individual patients, ensuring more personalised and effective care.

Table 1. Primary sources of RWD

Primary source	Details
Electronic health records (EHRs)	EHRs are digital versions of patients' medical records and are maintained by healthcare providers. These records include comprehensive data about patients' demographics, diagnoses, treatment histories, lab results, and prescriptions. EHRs provide a continuous, longitudinal view of patient care, making them one of the most valuable sources of RWD.[4] By analysing EHR data, researchers can track disease progression, evaluate treatment efficacy, and monitor patient outcomes across different healthcare settings.
Health insurance claims data	Health insurance claims data offers a wealth of information regarding healthcare service utilisation. This includes records of medical services such as hospitalisations, doctor visits, prescriptions, and diagnostic procedures, as well as reimbursement details.[1] Claims data is instrumental in studying healthcare patterns, understanding the cost of care, and assessing the impact of different treatment strategies on patient outcomes.
Patient registries	Patient registries are specialised databases that collect detailed information about individuals diagnosed with specific diseases or conditions. These registries serve as critical tools for tracking disease prevalence, monitoring treatment patterns, and conducting long-term outcomes research.[3] They are particularly valuable in understanding rare diseases or conditions that may not be adequately represented in clinical trials.
Wearable devices and mobile health applications	With the rise of digital health technologies, wearable devices (e.g., fitness trackers, smartwatches) and mobile health apps have become popular sources of real-time health data. These devices can measure physical activity, heart rate, sleep patterns, blood pressure, and more.[2] The data generated by wearables and apps provide continuous monitoring, offering valuable insights into patient behaviours and health outcomes in the real world.
Pharmacy records	Pharmacy records contain information about medications that are prescribed, dispensed, and used by patients. These records offer insights into medication adherence, treatment patterns, and therapeutic outcomes.[5] Analysing pharmacy data can help assess the effectiveness of drug therapies, identify trends in medication usage, and improve strategies for managing chronic diseases.

Image: Freepik

Table 2. Key difference in the data's stage of processing and its application[4]

	RWD	RWE
Raw data vs. analysed evidence	RWD refers to the raw, unprocessed data that is collected from various healthcare settings. It represents the factual information about healthcare practices and patient experiences.	RWE is the analysed, interpreted evidence derived from this raw data.[2] RWE provides insights that are actionable and can be used to inform clinical decision-making and policy formulation.
Data vs. insight	RWD consists of diverse, unstructured, and often complex datasets. It can be difficult to extract meaningful information from this raw data without appropriate analysis.	RWE, on the other hand, simplifies this complexity by distilling the data into actionable insights that healthcare professionals, policymakers, and researchers can use to improve care and outcomes.[1]
Purpose and application	The purpose of RWD is to collect comprehensive data about real-world healthcare practices, patient behaviours, and treatment outcomes.	The goal of RWE is to apply this data to improve clinical outcomes, guide policy decisions, and inform pharmaceutical development.
Complexity	RWD is inherently more complex due to its heterogeneity. It comes from various sources with varying formats and standards.	RWE simplifies this complexity by providing evidence that is directly applicable to real-world healthcare situations, making it easier for stakeholders to interpret and use in decision-making.[5]

Image: Freepik

5. **Continuous feedback loop**: One of the distinguishing features of RWE is that it is not static. As new data is continuously collected and analysed, the evidence base evolves.[7] This ongoing process allows healthcare systems to refine their practices, adapt to emerging trends, and improve patient care over time.

Real-world applications of RWD and RWE

The applications of RWD and RWE are diverse and far-reaching. They have the potential to revolutionise many aspects of healthcare delivery, from improving patient outcomes to guiding public health policy. Below are some key areas where RWD and RWE are making an impact:[3]

1. **Medication safety**: RWD, such as health insurance claims data and EHRs, can be used to identify potential adverse drug reactions (ADRs). This type of data provides a comprehensive view of patient treatment patterns and outcomes, helping researchers detect safety signals that may not be evident in controlled clinical trials.[1] RWE derived from this data can confirm the safety profile of medications, guide safety monitoring, and influence regulatory decisions regarding drug approval and use.
2. **Chronic disease management**: Chronic diseases such as diabetes, heart disease, and asthma are prevalent worldwide, and managing these conditions effectively requires ongoing monitoring and tailored treatment plans. RWD collected from EHRs and patient registries allows healthcare providers to track the long-term outcomes of patients with chronic diseases.[4] RWE generated from this data can help assess the effectiveness of various treatment regimens, refine clinical guidelines, and improve care management strategies.
3. **Pharmaceutical development**: RWD plays a critical role in pharmaceutical development by providing insights into disease prevalence, patient demographics, and treatment patterns. This information can be used to inform the design of clinical trials, making them more reflective of the real-world patient population.[6] Additionally, RWE is used in post-marketing surveillance to monitor the long-term safety and efficacy of drugs once they are available to the public.
4. **Healthcare policy and disparities**: RWD is also a powerful tool for identifying disparities in healthcare access and outcomes. By analysing data from different populations, policymakers can gain insights into the social, economic, and geographic factors that contribute to health inequities.[3] RWE can be used to shape policies aimed at reducing healthcare disparities, improving access to care, and promoting health equity.

Discussion and conclusion

The integration of RWD and RWE into healthcare practice is a monumental shift that holds the potential to transform how healthcare systems operate.[1] RWD provides a comprehensive view of real-world patient care, while RWE offers actionable insights that can guide decision-making at every level of the healthcare system. Together, these tools enable a more personalised, data-driven approach to healthcare, with the ability to improve patient outcomes, streamline clinical practices, and inform policy decisions.[2]

As healthcare continues to evolve, the role of RWD and RWE will become increasingly central in optimising patient care and shaping evidence-based healthcare policies. By continually collecting and analysing RWD, healthcare providers, researchers, and policymakers can ensure that healthcare practices are aligned with the realities of patient care, ultimately leading to better health outcomes on a global scale. The future of healthcare is one that integrates RWD and RWE into every aspect of decision-making, driving innovation, improving efficiency, and promoting health equity.

Disclosures and conflicts of interest

The author declares no conflicts of interest.

References

1. US FDA. Framework for FDA's Real-World Evidence Program. 2018. Available from: https://www.fda.gov/media/120060/download
2. Sherman RE, Anderson SA, Dal Pan GJ, et al. (2016). Real-world evidence – What is it and what can it tell us? N Engl J Med. 2016;375(23):2293-97. doi:10.1056/NEJMsb1609216
3. Garrison LP, Neumann PJ, Erickson P, et al. (2007). Using real-world data for coverage and payment decisions: The ISPOR Real-World Data Task Force report. Value Health. 2007;10(5):326-35. doi:10.1111/j.1524-4733.2007.00186.x3.
4. Corrigan-Curay J, Sacks L, Woodcock J, et al. (2018). Real-world evidence and real-world data for evaluating drug safety and effectiveness. JAMA. 2018;320(9):867-8. doi:10.1001/jama.2018.10136
5. Berger ML, Sox H, Willke R, et al. Good practices for real-world data studies of treatment and/or comparative effectiveness: Recommendations from the ISPOR Real-World Data Task Force on real-world evidence in health care decision making. Pharmacoepidemiol Drug Saf. 2017;26(9):1033-39. doi:10.1002/pds.4297
6. Miksad RA, Abernethy AP. Harnessing the power of real-world evidence: A checklist to ensure regulatory-grade data quality. Clin Pharmacol Ther. 20182018;103(2):202-5. doi:10.1002/cpt.946
7. Haris G, Acha A, Liu S, et al. Real-world evidence (RWE) to support regulatory submissions: Landscape assessment and review [Abstract HPR14]. Value Health. 2023;26(6):S213. doi:10.1016/j.jval.2023.03.1161

Author information

Andrew Balkin has been a medical writer for more than 20 years, and recently co-founded Ambermed Medical Writing and Consulting.

Real-world data and evidence: A European regulatory perspective

Somsuvro Basu
P95 Clinical and Epidemiology Services, Leuven, Belgium

doi: 10.56012/rhcw1179

Correspondence to:
Somsuvro Basu
basu.somsuvro@gmail.com

Abstract

The European Medicines Agency (EMA) has increasingly recognised the value of real-world data (RWD) and real-world evidence (RWE) in this ever-changing regulatory landscape. Regulatory decisions have been traditionally based primarily on data from randomised controlled trials. However, RWD – information collected from everyday clinical settings such as patient registries, health records, and insurance claims – has emerged as a complementary resource. The EMA views RWE as a key tool for enhancing drug development, supporting adaptive licensing, and providing a more accurate assessment of the safety and effectiveness of medicines once they are on the market. The agency emphasises a structured approach to incorporating RWD through frameworks, pilot programmes, and data standardisation initiatives. Through various initiatives, including the DARWIN EU® initiative and its Big Data Steering Group, the EMA is working to ensure the scientific validity and regulatory acceptability of RWE across the medicinal product lifecycle.

Image: Freepik

Introduction

The European Medicines Agency (EMA) has been increasingly incorporating real-world evidence (RWE) into its regulatory framework to support decision-making across the lifecycle of medicines. This integration reflects the growing recognition of RWE's potential to complement traditional clinical trial data, particularly in addressing evidence gaps, monitoring safety, and informing post-approval assessments.[1] EMA's approach to RWE is structured around several key foundations, including the generation of fit-for-purpose data, collaboration with stakeholders, and the development of robust methodologies to ensure the quality and reliability of RWE.[2]

In their report "EMA Regulatory Science to 2025: Strategic reflection," EMA proposed "Driving collaborative RWE generation to improve the scientific quality of evaluations" as one of the five strategic goals in human medicine.[3] EMA's evolving strategy emphasises the need for continuous dialogue with stakeholders to optimise the use of RWE in regulatory evaluations.[4] This shift aims to enhance the overall benefit-risk assessment of medicinal products. Moreover, EMA's commitment to integrating RWE signifies a paradigm shift in regulatory science, fostering a more comprehensive understanding of medicines' real-world performance and safety profiles.[4] This progressive approach acknowledges some limitations of traditional randomised controlled trials (e.g., small and highly selected sample size) and seeks to leverage the growing availability of digital health data for more effective regulatory decisions.[5] This article explores the technical methods, practical applications, and specific case studies of RWE integration into EMA's regulatory framework, highlighting the challenges and future directions in this evolving field.

Table 1. OPTIMAL (OPerational-TechnIcal-MethodologicAL) framework for real-world evidence[a]

Pillar	Focus area	Key components
Operational	Concerns the **quality and governance** of real-world data and data sources	• Data provenance, access, and governance structures • Completeness, consistency, accuracy, traceability, and sustainability of data collection • Data sharing agreements aligned with the General Data Protection • Regulation and national legislation; use of ENCePP Code of Conduct • Transparent documentation of data source policies and collaboration models
Technical	Relates to **data structure, interoperability, and management systems**	• Use of standardised terminologies and coding systems • Interoperability across systems and potential for data linkage (e.g., registries, electronic health records, claims) • Mapping to Common Data Models, including validation • Quality assurance/control procedures, including internal/external audits • Benchmarking against external data sources • Capture of critical time elements and consistent recording practices • EMA qualification procedure for data source
Methodological	Focuses on the **scientific rigour** of study design and analysis of real-world evidence	• Use of appropriate, validated study designs for regulatory purposes • Identification and control of confounding factors and biases • Documentation of feasibility analysis • Study protocol registration and transparent reporting of results • Use of best practices in epidemiology/statistics • Seeking EMA Scientific Advice for study protocol evaluation

[a]Adapted table from *Cave et al. (2019)*[6]

Abbreviations: EMA, European Medicines Agency; ENCePP, European Network for Centres of Pharmacoepidemiology and Pharmacovigilance.

EMA's regulatory framework and actions on RWE

EMA has taken a proactive approach to integrating RWE into its regulatory decision-making framework. This includes strategic initiatives, infrastructure development, pilot programmes, and stakeholder collaboration aimed at fostering trust in the utility of RWD and RWE.

Key pillars of the RWE framework

EMA's approach to RWE is structured around three central pillars – operational, technical, and methodological (Table 1), forming the OPTIMAL framework,[6] which could collectively aim to maximise the potential of RWE in regulatory decision-making.

EMA has identified three major pathways for generating RWE: the Data Analysis and Real World Interrogation Network (DARWIN EU®), in-house electronic health databases, and studies commissioned via EMA framework contracts.[2] Building on these and the key pillars, EMA has introduced several initiatives aimed at bolstering its commitment to RWE.

The Big Data Steering Group

In 2020, EMA and the Heads of Medicines Agencies established the Big Data Steering Group (BDSG) to enhance the capacity of the European regulatory network to use big data, including RWD. This group provides strategic direction on priorities, such as data quality, analytics capabilities, and RWE acceptability. They aim to advance the integration of big data into the regulatory framework to improve public health outcomes.[7] BDSG is currently exploring the use of mobile health (mHealth) data, generated from devices like smartphones and wearables, to provide detailed patient information (e.g., heart rate and sleep quality) for regulatory decisions. While mHealth data shows promise, challenges such as data quality and privacy must be addressed.[8]

DARWIN EU®

Launched in 2022, DARWIN EU® is a key initiative enabling the EMA to access and analyse RWD from across Europe. By establishing a network of data partners across the EU, DARWIN EU® facilitates timely and robust analyses, enabling regulators to assess the safety and effectiveness of medicinal products more efficiently.[9] This network aims for approximately 40 partners by February 2026 and currently provides access to anonymised health data from around 180 million patients across Europe.[10]

DARWIN EU® assists in:

- Supporting regulatory assessments, such as risk-benefit evaluations, safety signals, and post-marketing surveillance.
- Enhancing the understanding of disease epidemiology and treatment patterns.
- Contributing to research and innovation by enabling high-quality studies at scale.

Guidance documents and scientific advice

The EMA provides specific guidance on the design, conduct, and reporting of RWE studies. For example, the "Guideline on Registry-Based Studies"[11] and the "Guideline on Good Pharmacovigilance Practices" outline expectations for the quality and transparency of RWE.[12]

Additionally, developers can seek scientific advice from the EMA on planned RWE studies.[2] This helps align study designs with regulatory

expectations early in the process, thereby increasing the likelihood that RWE will be considered acceptable during regulatory review.

Pilot programmes and case studies

The EMA has conducted various pilots to test the regulatory acceptability of RWE. For instance, a pilot programme involving post-authorisation safety studies demonstrated how RWE could supplement traditional data sources in assessing long-term safety.[13] Another pilot programme used patient registries to evaluate treatment effectiveness in rare diseases, supporting product label extensions.[14,15]

Through these pilots, EMA has demonstrated that well-conducted RWE studies can provide reliable and actionable evidence for regulatory purposes.

Technical methods for RWE integration

EMA has developed and refined various technical methods to incorporate RWE into its regulatory processes. These methods are designed to address the complexities of RWD, ensuring its quality, reliability, and relevance for decision-making.

Study design and data sources

RWE studies often involve retrospective cohort studies, case-control studies, and observational research. EMA emphasises the importance of using fit-for-purpose data sources, such as electronic health records, claims databases, and patient registries.[16] For instance, EMA's pilot programme identified 61 research topics for RWE generation, with a focus on medicine safety, clinical trial design, drug utilisation, and disease epidemiology, and each necessitates fit-for-purpose data sources relevant to the regulatory research question.[17] Additionally, EMA has stressed the importance of early interactions with stakeholders, such as marketing authorisation holders, to better understand research questions and optimise RWE generation.[17,18] Similarly, the use of external controls derived from RWD has been explored to contextualise outcomes from uncontrolled trials, though this requires careful consideration of eligibility criteria, temporality, and population representation.[19]

Analytical methods

The use of advanced analytical techniques, such as propensity score matching and instrumental variable analysis, can help address confounding and bias in RWE studies.[20] EMA also advocates for transparency in study design and analysis, with a focus on reproducibility and robustness.[21] EMA recommends sensitivity analyses to be considered to assess the reliability of findings in the context of uncertainty and potential bias.[2,19,22,23]

> By establishing a network of data partners across the EU, DARWIN EU® facilitates timely and robust analyses, enabling regulators to assess the safety and effectiveness of medicinal products more efficiently.

Data quality and standardisation

Ensuring the quality of RWD is a critical challenge. EMA has emphasised the need for standardised data collection and reporting practices. Initiatives such as the European Health Data and Evidence Network (EHDEN)[24] and the GetReal Institute[25] aim to improve data interoperability and facilitate the use of RWE in regulatory and Health Technology Assessment (HTA) processes.[26,27]

Practical applications of RWE in the EU regulatory framework

Practical applications of RWE have emerged across various stages of the medicinal product lifecycle, from pre-approval to post-marketing surveillance.

Regulatory decision-making

RWE is increasingly used to support regulatory decisions, particularly in cases where clinical trial data is limited. For example, RWE has been instrumental in evaluating the safety and effectiveness of orphan medicinal products to overcome the unique challenges of rare disease drug development.[28] The EMA has also used RWE to inform decisions on drug utilisation and clinical management, particularly in the context of the COVID-19 pandemic.[17]

In regulatory submissions, RWE has been used in the context of marketing authorisation applications (MAAs) and extensions of indication. A review of MAAs submitted to the EMA in 2018 and 2019 revealed that 40% of these applications included RWE, primarily derived from registries and hospital data, to support safety and efficacy assessments.[29] Similarly, RWE has been used to inform post-marketing safety evaluations, particularly for rare adverse events and subgroup analyses.[30] EMA has also recognised the value of RWE in evaluating the effectiveness of risk minimisation measures and understanding product usage and misuse.[30]

> EMA's pilot programme identified 61 research topics for RWE generation, with a focus on medicine safety, clinical trial design, drug utilisation, and disease epidemiology.

Health technology assessment

RWE plays a pivotal role in HTA processes, particularly in evaluating the real-world effectiveness and cost-effectiveness of treatments. HTA bodies such as the National Institute for Health and Care Excellence (NICE) and the Haute Autorité de Santé (HAS) have increasingly accepted RWE to inform reimbursement decisions, though challenges remain in terms of data quality and generalisability.[22,31] Moreover, the collaborative work between EMA and EUnetHTA (European Network for Health Technology Assessment) highlights the importance of using extrapolation (inferring data from adults to children, aiding regulatory decision-making) in paediatric drug development to address ethical concerns and support robust evidence generation, ultimately aiding in benefit/risk considerations for regulatory authorities and HTA bodies.[32]

Post-approval surveillance

RWE is widely used for post-marketing surveillance, enabling the identification of safety signals and adverse events in real-world settings. The EMA's DARWIN EU® initiative, a network for RWD analysis, has been instrumental in monitoring the safety of approved medicinal products.[27]

Healthcare resource use

RWE is increasingly used to assess the economic impact of healthcare interventions. By providing insights into treatment patterns, resource utilisation, and long-term outcomes, RWE supports value-based decision-making and resource allocation.[22,33]

Image: Freepik

Case studies: RWE in action across the EU

Several case studies illustrate the successful integration of RWE into EMA's regulatory framework.

Orphan medicinal products

RWE has been particularly valuable in the approval of orphan medicinal products, where clinical trial data is often limited. For example, the approval of abaloparatide for osteoporosis relied on RWE to address gaps in clinical trial data, demonstrating the effectiveness of RWE in supporting regulatory decisions for rare diseases.[34]

Oncology medicines

In oncology, RWE has been used to evaluate the real-world effectiveness of cancer treatments. A review of oncology-targeted therapies approved between 2018 and 2022 indicated that RWE contributed to regulatory decisions in 21% of cases, demonstrating its role in addressing evidence gaps and enhancing understanding of treatment outcomes.[35] A case study on the approval of a novel oncology medicine highlighted the role of RWE in bridging the gap between clinical trial results and real-world outcomes, facilitating regulatory and HTA decisions.[36]

Rare diseases and registry data

The SATURN initiative demonstrates the feasibility of using existing registries to collect RWE for rare diseases. By leveraging data from the Registry of Osteogenesis Imperfecta, SATURN has provided valuable insights into treatment practices and outcomes, supporting regulatory and HTA decision-making.[37]

Challenges in RWE integration

Despite its potential, integrating RWE into the EMA's regulatory framework faces several challenges.

Data heterogeneity and quality

The heterogeneity of RWD sources across EU member states poses significant challenges. Differences in healthcare systems, data collection practices, and privacy regulations complicate the aggregation and analysis of RWD.[6,33]

A review of oncology-targeted therapies approved between 2018 and 2022 indicated that RWE contributed to regulatory decisions in 21% of cases, demonstrating its role in addressing evidence gaps and enhancing understanding of treatment outcomes.

Regulatory harmonisation

The lack of harmonised guidelines for RWE use across EU member states and HTA bodies remains a barrier. While the EMA has made progress in developing RWE-specific guidance, inconsistencies in terminology and methodological preferences persist.[26,38]

Ethical and privacy considerations

The use of RWD raises ethical and privacy concerns, particularly concerning patient confidentiality and data protection. The implementation of the General Data Protection Regulation in the EU has introduced additional complexities in RWD utilisation.[22,33]

Image: Freepik

Future directions for RWE in the EU

To fully realise the potential of RWE, the EMA and other stakeholders must address existing challenges and invest in initiatives that enhance the quality, accessibility, and harmonisation of RWD.

Harmonising RWE guidelines

The development of harmonised guidelines for RWE use across EU member states and HTA bodies is essential. A public-private partnership, the Integration of heterogeneous Data and Evidence towards Regulatory & HTA Acceptance (IDERHA) project,[39] aims to align RWE requirements and reduce fragmentation in regulatory and HTA processes.[38,40] IDERHA, launched in April 2023, aims to apply artificial intelligence (AI) and machine learning (ML) to link and analyse diverse health data for early detection of lung cancer and improved quality of life for those affected.[39]

Enhancing data infrastructure

Investing in robust data infrastructure is critical to overcoming the limitations of RWD. Under the IDERHA project, initiatives such as the European Health Data Space (EHDS) aim to improve data interoperability and facilitate cross-border data sharing.[6,41] EHDS is designed to enable individuals to access and manage their health data across the EU. This initiative includes the primary use of data (EHDS1; MyHealth@EU) for healthcare delivery and decision-making, and the secondary use of data (EHDS2; HealthData@EU) for research, innovation, policy-making, and regulatory purposes.[42]

Promoting innovation and collaboration

The integration of advanced technologies, such as AI and ML, holds promise for enhancing RWE generation and analysis. Collaboration among regulators, industry stakeholders, and academia will be key to driving innovation and addressing methodological challenges.[22,43]

Building on existing initiatives

The EMA's Network Strategy to 2025 aims to leverage RWE further in regulatory processes, addressing critical research questions and enhancing medicinal product evaluations.[44] Building on that, EMA's Network Strategy 2028 plans to further leverage RWE with a strengthening contribution from digital transformation and AI.[45]

Conclusion

EMA is at the forefront of integrating RWE into the regulatory framework. Through strategic initiatives like DARWIN EU®, robust guidance documents, and collaborative stakeholder engagement, EMA is enabling the generation and use of high-quality, regulatory-grade RWE. Despite challenges in data quality, methodology, and interoperability, the outlook is promising. Continued investment in data infrastructure, methodological innovation, and regulatory alignment will be key to unlocking the full potential of RWE. As RWE becomes increasingly embedded in decision-making processes, it promises to enhance the evaluation of medicinal products, improve patient outcomes, and support more dynamic and responsive regulatory practices.

Acknowledgements

The author would like to thank Ana Goios (P95 Clinical and Epidemiology Services, Leuven, Belgium) for reviewing the manuscript.

Disclaimers

The opinions expressed in this article are the author's own and not necessarily shared by his employer, the EMWA, or the EMA.

Disclosures and conflicts of interest

The author is employed by P95 Clinical and Epidemiology Services (Leuven, Belgium), but this work is independent of his employment.

In addition to PubMed, SciSpace and ChatGPT were utilised for the preliminary literature search.

References

1. EMA. Use of real-world evidence in regulatory decision making – EMA publishes review of its studies. 2023 [cited 2025 July 17]. Available from: https://www.ema.europa.eu/en/news/use-real-world-evidence-regulatory-decision-making-ema-publishes-review-its-studies
2. EMA. Real-world evidence provided by EMA. 2024 [cited 2025 July 17]. Available from: https://www.ema.europa.eu/en/documents/other/guide-real-world-evidence-provided-ema-support-regulatory-decision-making_en.pdf
3. EMA. EMA Regulatory Science to 2025 Strategic reflection. 2019 [cited 2025 July 17]. Available from:

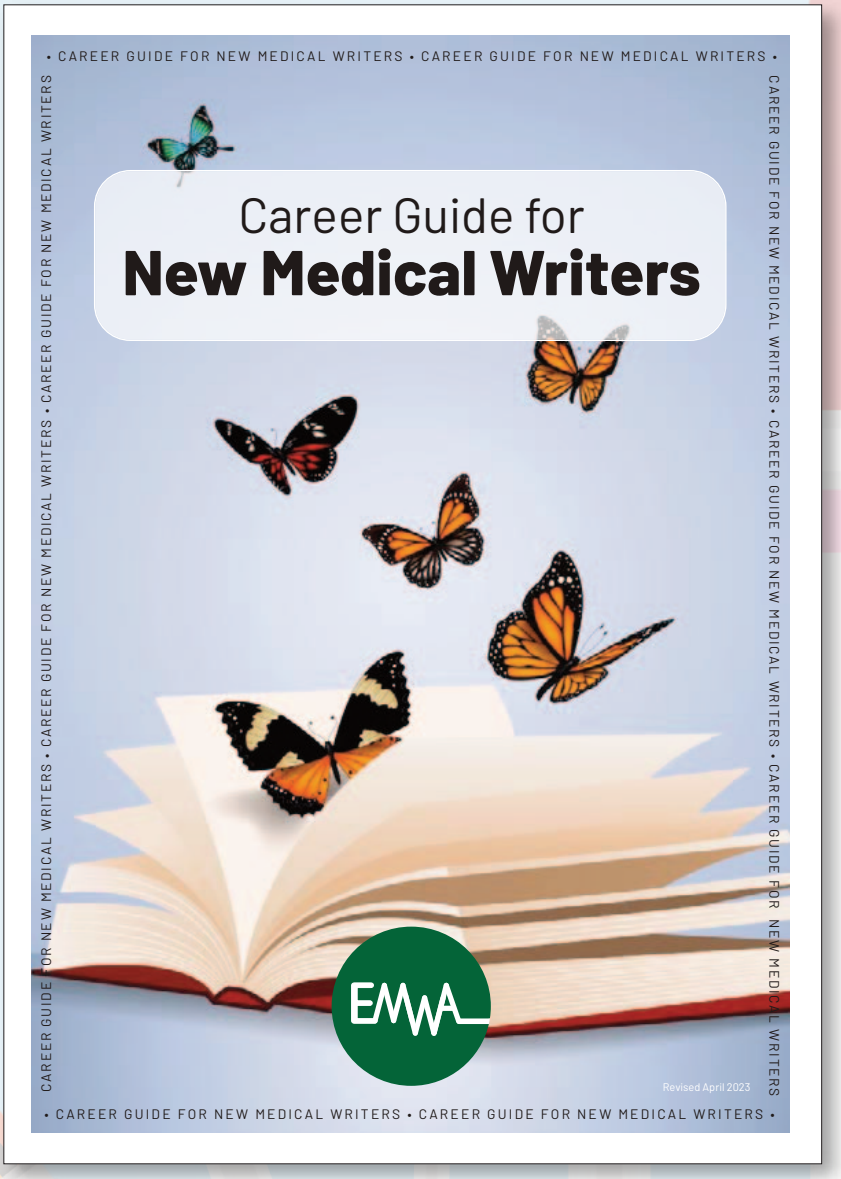
https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/ema-regulatory-science-2025-strategic-reflection_en.pdf
4. Bakker E, Plueschke K, Jonker CJ, et al. Contribution of real-world evidence in European Medicines Agency's regulatory decision making. Clin Pharmacol Ther. 2023;113(1):135-151. doi.10.1002/cpt.2766
5. Maison P, Zureik M, Hivert V, et al. Real-world evidence (RWE): A challenge for regulatory agencies discussion of the RWE conference with the network of the European Medicine Agencies, patients, and experts [published correction appears in Front Pharmacol. 2025 Apr 07;16:1601645 doi:10.3389/fphar.2025.1601645.]. Front Pharmacol. 2022;13:969091. doi:10.3389/fphar.2022.969091.
6. Cave A, Kurz X, Arlett P. Real-world data for regulatory decision making: Challenges and possible solutions for Europe. Clin Pharmacol Ther. 2019;106(1):36-9 doi:10.1002/cpt.142610.1002/cpt.1426
7. EMA. Big Data Workplan 2023-2025: HMA/EMA joint Big Data Steering Group. 2024 [cited 2025 July 18]. Available from: https://www.ema.europa.eu/en/documents/work-programme/workplan-2023-2025-hma-ema-joint-big-data-steering-group_en.pdf
8. Heads of Medicines Agencies (HMA). mHealth Data for Real World Evidence in Regulatory Decision Making. An expert review report for the HMA/EMA Big Data Steering Group-2024. 2024 [cited 2025 July 17]. Available from: https://www.ema.europa.eu/system/files/documents/report/mhealth-data-regulatory-decision-making_expert-review-report_en.pdf
9. EMA. Data Analysis and Real World Interrogation Network (DARWIN EU). 2025 [cited 2025 May 16]. Available from: https://www.ema.europa.eu/en/about-us/how-we-work/data-regulation-big-data-other-sources/real-world-evidence/data-analysis-real-world-interrogation-network-darwin-eu
10. DARWIN-EU. Home Page. 2025 [cited 2025 July 17]. Available from: https://www.darwin-eu.org/
11. EMA. Committee for Human Medicinal Products (CHMP) Guideline on registry-based studies. 2021 [cited 2025 July 17]. Available from: https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-registry-based-studies_en.pdf
12. EMA. Good pharmacovigilance practices (GVP). 1995–2025 [cited 2025 July 17]. Available from: https://www.ema.europa.eu/en/human-regulatory-overview/post-authorisation/pharmacovigilance-post-authorisation/good-pharmacovigilance-practices-gvp
13. EMA. Post-authorisation safety studies (PASS).1995–2025 [cited 2025 July 17]. Available from: https://www.ema.europa.eu/en/human-regulatory-overview/post-authorisation/pharmacovigilance-post-authorisation/post-authorisation-safety-studies-pass
14. EMA. Patient registries. 1995 – 2025 [cited 2025 July 17]. Available from: https://www.ema.europa.eu/en/human-regulatory-overview/post-authorisation/patient-registries#related-content-17663
15. Jonker CJ, Bakker E, Kurz X, et al. Contribution of patient registries to regulatory decision making on rare diseases medicinal products in Europe. Front Pharmacol. 2022;13:924648. Published 2022 Aug 4. doi:10.3389/fphar.2022.924648.
16. EMA. Report from "Multi-stakeholder workshop on Real World Data (RWD) quality and Real World Evidence (RWE) use". 2023 [cited 2025 July 17]. Available from: https://www.ema.europa.eu/en/documents/report/report-multi-stakeholder-workshop-real-world-data-rwd-quality-and-real-world-evidence-rwe-use_en.pdf
17. Prilla S, Groeneveld S, Pacurariu A, et al. Real-world evidence to support EU regulatory decision making – results from a pilot of regulatory use cases. Clin Pharmacol Ther. 2024;116(5):1188-1197. doi:10.1002/cpt.3355
18. EMA. Real-world evidence framework to support EU regulatory decision-making - 2nd report on the experience gained with regulator-led studies from February 2023 to February 2024. 2024 [cited 2025 July 17]. Available from: https://www.ema.europa.eu/en/documents/report/real-world-evidence-framework-support-eu-regulatory-decision-making-2nd-report-experience-gained-regulator-led-studies-february-2023-february-2024_en.pdf
19. Curtis LH, Sola-Morales O, Heidt J, et al. Regulatory and HTA considerations for development of real-world data derived external controls. Clin Pharmacol Ther. 2023;114(2):303-15. doi:10.1002/cpt.2913
20. Li Q, Lin J, Chi A, Davies S. Practical considerations of utilizing propensity score methods in clinical development using real-world and historical data. Contemp Clin Trials. 2020;97:106123. doi:10.1016/j.cct.2020.106123
21. EMA. Transparency. 1995 –2025 [cited 2025 July 17]. Available from: https://www.ema.europa.eu/en/about-us/how-we-work/transparency#clinical-trials-12744
22. Bhatia N. Harnessing real-world evidence in pharmacoeconomics: A comprehensive review. Open Health. 2024;5(1):20230048. doi:10.1515/ohe-2023-0048
23. EMA. Reflection paper on use of real-world data in non-interventional studies to generate real-world evidence for regulatory purposes. 2025 [cited 2025 July 17]. Available from: 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
24. European data. European Health Data and Evidence Network (EHDEN): Shaping the future of health data in Europe. 2024 [cited 2025 July 17]. Available from: https://data.europa.eu/en/news-events/news/european-health-data-and-evidence-network-ehden-shaping-future-health-data-europe
25. GetReal Institute. 2025 [cited 2025 May 16]. Available from: https://getreal-institute.org/
26. Sarri G, Hernandez LG. The maze of real-world evidence frameworks: from a desert to a jungle! An environmental scan and comparison across regulatory and health technology assessment agencies. J Comp Eff Res. 2024;13(9):e240061. doi:10.57264/cer-2024-0061

27. Claire R, Elvidge J, Hanif S, et al. Advancing the use of real world evidence in health technology assessment: insights from a multi-stakeholder workshop. Front Pharmacol. 2024;14:1289365. Published 2024 Jan 12. doi:10.3389/fphar.2023.1289365
28. Naumann-Winter F, Wolter F, Hermes U, et al. Licensing of orphan medicinal products – Use of real-world data and other external data on efficacy aspects in marketing authorization applications concluded at the European Medicines Agency between 2019 and 2021. Front Pharmacol. doi:10.3389/fphar.2022.920336
29. Flynn R, Plueschke K, Quinten C, et al. Marketing authorization applications made to the European Medicines Agency in 2018–2019: What was the contribution of real-world evidence? Clin Pharmacol Ther. 2022;111(1):90-97. doi:10.1002/cpt.2461
30. Brown JP, Wing K, Evans SJ, et al. Use of real-world evidence in postmarketing medicines regulation in the European Union: a systematic assessment of European Medicines Agency referrals 2013–2017. BMJ Open. 2019;9(10):e028133. doi:10.1136/bmjopen-2018-028133
31. Al-khayat Z, Franzen N, Retèl VP, et al. PP37 guidance on using hospital-based real-world evidence in health technology assessments for oncology. Int J Technol Assess Health Care. 2024;40(S1):S69. doi:10.1017/S0266462324002095
32. Karres D, Pino-Barrio MJ, Benchetrit S, et al. Evidence generation throughout paediatric medicines life cycle: findings from collaborative work between European Medicines Agency (EMA) and EUnetHTA on use of extrapolation. Br J Pharmacol. 2025;182(3):484-494. doi:10.1111/bph.17396
33. Zisis K, Pavi E, Geitona M, et al. Real-world data: a comprehensive literature review on the barriers, challenges, and opportunities associated with their inclusion in the health technology assessment process. J Pharm Pharm Sci. 2024;27:12302. doi:10.3389/jpps.2024.12302
34. Davenport C, Gravel P, Wang Y, et al. Real-world evidence to support the registration of a new osteoporosis medicinal product in Europe. Ther Innov Regul Sci. 2024;58(3):505-18. doi:10.1007/s43441-024-00616-7
35. Derksen JW, Martins-Branco D, Valachis A, et al. Real-world evidence reported for clinical efficacy evaluation in European Public Assessment Reports of authorised targeted therapies for solid malignancies: a comprehensive review (2018-2022). ESMO Real World Data and Digital Oncology. 2024;4:100039. doi:10.1016/j.esmorw.2024.100039
36. Zong J, Rojubally A, Pan X, et al. A review and comparative case study analysis of real-world evidence in European regulatory and health technology assessment decision making for oncology medicines. Value Health. 2025;28(1):31-41. doi:10.1016/j.jval.2024.09.007
37. Sangiorgi L, Boarini M, Westerheim I, et al. Project SATURN – a real-world evidence data collaboration with existing European datasets in Osteogenesis Imperfecta to support future therapies. Orphanet J Rare Dis. 2024;19(1):184. doi:10.1186/s13023-024-03185-y
38. Cresswell K, Claire R, Avsar TS, et al. PD137 Reviewing the health technology assessment and regulatory policy landscape on acceptability standards for real-world evidence – Initial findings. Int J Technol Assess Health Care.2024;40(S1):S147-S147. doi:10.1017/S026646232400374X
39. IDERHA. 2025 [cited 2025 July 17]. Available from: https://www.iderha.org/
40. Thokagevistk K, Coppo C, Rey L, et al. Real-world evidence to reinforce clinical trial evidence in health technology assessment: A critical review of real-world evidence requirements from seven countries and recommendations to improve acceptance. J Mark Access Health Policy. 2024;12(2):105–17. doi:10.3390/jmahp1202000909
41. Li M, Chen S, Lai Y, et al. Integrating real-world evidence in the regulatory decision-making process: A systematic analysis of experiences in the US, EU, and China using a logic model. Front Med (Lausanne). 2021;8:669509. doi:10.3389/fmed.2021.669509
42. Hussein R, Balaur I, Burmann A, et al. Getting ready for the European Health Data Space (EHDS): IDERHA's plan to align with the latest EHDS requirements for the secondary use of health data. Open Res Eur. 2024;4:160. doi:10.12688/openreseurope.18179.1
43. Jansen MS, Dekkers OM, le Cessie S, et al. Real-world evidence to inform regulatory decision making: A scoping review. Clin Pharmacol Ther. 2024;115(6):1269-76. https://doi.org/10.1002/cpt.3218
44. EMA. European medicines agencies network strategy to 2025 - Protecting public health at a time of rapid change. 2020 [cited 2025 July 17]. Available from: https://www.ema.europa.eu/system/files/documents/report/eman_strategy_2025_en.pdf
45. EMA. Seizing opportunities in a changing medicines landscape. 2024 [cited 2025 July 17]. Available from: https://www.ema.europa.eu/en/news/seizing-opportunities-changing-medicines-landscape

Author information

Somsuvro Basu (Som) is a Senior Medical Writer at P95 Clinical and Epidemiology Services, Leuven, Belgium. Som has doctoral (University of South Bohemia, Biology Centre – Czech Academy of Sciences, Czech Republic) and postdoctoral (Philipps-University Marburg, Germany) training in molecular cell biology (parasitology and human mitochondrial biology).

Som previously worked as a Science Communication Officer at the Central European Institute of Technology. Parallelly, he acted as the EU-LIFE Science Communications working group co-chair. He has been an active EMWA member since 2017 and volunteered as the EMWA Honorary Secretary (2021-2025) and is on the editorial board of EMWA's *Medical Writing* journal.



Career Guide for New Medical Writers

EMWA's Getting into Medical Writing group has created an updated *Career Guide for New Medical Writers*, which is available on the EMWA website. If you're new to medical writing, it's a useful resource that will help you take your first steps on this rewarding career path.

You can email us at gettingintoMW@emwa.org.

From data to impact: Exploring the evolution of real-world evidence at the FDA

Shadaab Ayub[1], **Scott Swain**[2]
[1] Parexel International, Hyderabad, India
[2] Parexel International LLC, Kansas, USA

doi: 10.56012/jdwg8880

Correspondence to
Shadaab Ayub
Shadaab.Ayub@parexel.com
Scott Swain
Scott.Swain@parexel.com

Abstract:
The healthcare industry has witnessed a significant shift towards the use of real-world data (RWD) and real-world evidence (RWE) in medical decision-making and treatment evaluation. This article explores the importance of RWE in healthcare decision-making and its potential to revolutionise drug development, regulatory processes, and patient care. We examined the US FDA's strategic vision for RWD and RWE, tracing the agency's evolving stance from early 2000s to present day. The FDA's framework for RWE, key guidance documents, and the establishment of the Center for Real-World Evidence Innovation are discussed. Case studies illustrate RWE's role in supporting regulatory decisions, including safety labelling changes and label expansions. We conclude by discussing the future of RWE, including its integration with artificial intelligence and its importance in evaluating cell and gene therapies, emphasising its transformative impact on healthcare innovation and regulatory processes.

Introduction

In recent years, the healthcare industry has undergone a significant shift in the methods used for medical decision-making and treatment evaluation. Leading this transformation is the growing acceptance of real-world data (RWD) and real-world evidence (RWE). This article explores the importance of RWE in healthcare decision-making and their potential to revolutionise drug development, regulatory processes, and patient care. In this context, the article will also examine the US FDA strategic vision for RWD and RWE.

RWD refers to all health-related information collected outside of traditional clinical trial settings and includes electronic health records, claims databases, as well as data from wearables, social media, and a variety of other sources; whereas RWE is the clinical evidence derived from the analysis of RWD. RWE provides valuable insights into real-world treatment patterns and unmet medical needs, helping to prioritise research efforts and allocate resources more effectively. Real-world research is particularly valuable for evaluating safety and effectiveness in rare or severe diseases, paediatric populations, or other patient groups often underrepresented in clinical trials, as it allows for an understanding of treatment effects in diverse patient populations receiving care in conventional settings. RWD can also support post-marketing surveillance, helping to identify emerging safety signals and long-term safety or to describe treatment outcomes among patients treated off-label. In regulatory approvals, RWE has been used to confirm the effectiveness of drugs approved under accelerated pathways, support label expansions, and provide contextual data for single-arm trials.

RWD can also support post-marketing surveillance, helping to identify emerging safety signals and long-term safety or to describe treatment outcomes among patients treated off-label.

RWE: FDA history and guidance

RWD and drug safety

The FDA has a long-standing history of utilising RWD and RWE for post-market safety monitoring. The FDA's evolving stance on RWE can be traced back to the early 2000s, with a notable milestone being the Prescription Drug User Fee Act (PDUFA) III in 2002. PDUFA III introduced user fees to expedite drug reviews and, in recognition that electronic healthcare data sources were becoming a valuable source for medical research, the act included provisions for considering observational data in regulatory decisions related to drug safety. This act laid the groundwork for the FDA's gradual integration of RWE into its regulatory framework.[1] The agency's Sentinel Initiative, launched in 2008, is a prime example of this approach. Sentinel uses a standard data model to combine various sources of insurance claims data and electronic health records (EHRs) to evaluate the safety of medical products after they have been approved and are in use in the general population.[2] The FDA has also used RWE in the context of rare diseases and paediatric populations, where conducting large-scale randomised clinical trials may be infeasible or unethical. For example, the agency has considered RWE from natural history studies and patient registries to support the approval of treatments for rare genetic disorders.[3]

Framework for RWE

The 21st Century Cures Act, signed into law in 2016, has significantly shaped the FDA's approach to evaluating and using RWE. This legislation mandated the FDA to establish a programme to evaluate the potential use of RWE to support approval of new indications for approved drugs or to satisfy post-approval study requirements.[4] Importantly, the Cures Act explicitly enabled the FDA to consider RWD in assessing drug effectiveness, expanding the potential applications of RWE beyond its traditional role in safety monitoring.

Key aspects of the FDA's RWE framework include:

- Ensuring transparency in the use of RWE for regulatory purposes.
- Defining appropriate use cases for RWE in regulatory decision-making.
- Establishing standards for data quality and reliability.
- Developing methodologies for analysing and interpreting RWE.

Image: Freepik

The framework emphasises the importance of data quality, study design, and analytical methods in generating reliable and relevant RWE. It also acknowledges the need for transparency in the use of RWE and the importance of stakeholder engagement in developing standards and best practices.[5]

Guidance documents

To support the implementation of its RWE framework, the FDA has issued several guidance documents that provide detailed recommendations on various aspects of RWE use in regulatory decision-making.

These guidance documents provide stakeholders with clear expectations and methodological considerations for using RWD and RWE in regulatory submissions. They address issues such as data collection and quality, study design, and analytical methods, helping to ensure that RWE is reliable, relevant, and appropriately integrated into the regulatory process.[6]

Key FDA guidance documents[6] include:

- **Use of Real-World Evidence to Support Regulatory Decision-Making for Medical Devices.**[7] This document focuses on how RWD and RWE can be used in regulatory decisions for medical devices. It covers aspects such as approval, labeling changes, and post-marketing surveillance.
- **Submitting Documents Using Real-World Data and Real-World Evidence to FDA for Drugs and Biologics.**[8] This guidance offers recommendations to sponsors on submitting documents containing RWD and RWE to the FDA. It focuses on the evaluation process for drugs and biological products using real-world information.
- **Real-World Data: Assessing Electronic Health Records and Medical Claims Data To Support Regulatory Decision-Making for Drug and Biological Products.**[9] This guidance offers recommendations on evaluating the relevance and reliability of EHR and medical claims data for generating RWE. It focuses on using these data sources to support regulatory decisions for drugs and biologics.
- **Data Standards for Drug and Biological Production Submissions Containing Real-World Data.**[10] This document establishes standardised data formats and requirements for RWD submissions in regulatory applications. It aims to ensure consistency and reliability in the evaluation process of drug and biological products.
- **Considerations for the Design and Conduct of Externally Controlled Trials for Drug and Biological Products.**[11] This document provides recommendations on designing, conducting, and analysing externally controlled trials. It focuses on using these trials to support regulatory decisions for drugs and biological products

A comprehensive list of FDA RWE guidance is currently available on the agency's website.[6]

The FDA's guidance emphasises the importance of data quality, including considerations such as data provenance, completeness, and accuracy. It also addresses methodological considerations for study design and analysis, recognising that the observational nature of much RWD requires careful attention to

potential biases and confounding factors.

Furthermore, the FDA's guidance documents highlight the Agency's commitment to an evolving healthcare system, where insights from routine clinical practice can inform regulatory decision-making and vice versa. This approach aims to create a more efficient and responsive regulatory process that can keep pace with rapid advancements in medical science and technology.

Additionally, the FDA Center for Drug Evaluation and Research (CDER) established the Center for Real-World Evidence Innovation (CCRI) in December 2024, marking a significant step in the Agency's commitment to advancing the use of RWE in regulatory decision-making. The objective of the CCRI is to promote and facilitate the use of RWE in drug development and regulatory processes, enhancing the efficiency and effectiveness of drug evaluation and approval. The CCRI was created in response to the increasing role of RWE in drug development and the need for a coordinated approach to its evaluation and integration in regulatory processes. This initiative is expected to significantly influence how RWE is utilised in drug approvals, potentially leading to more efficient and patient-centric regulatory decisions. The CDER CCRI insights are expected to inform future FDA policies and guidelines on RWE use, providing clearer direction to regulators and pharmaceutical companies regarding the design and analysis of fit-for-purpose RWE studies intended to support regulatory decision-making.[12]

> The FDA's guidance emphasises the importance of data quality, including considerations such as data provenance, completeness, and accuracy.

Case studies: RWE in decision making

The FDA has increasingly recognised the value of RWE in supporting faster and more informed regulatory decisions. RWE can provide additional evidence to support drug approvals, label expansions, or post-market safety monitoring.

Case study (safety labelling): Fluroquinolones

RWD data studies have played a crucial role in supporting labelling changes for fluoroquinolones regarding various safety issues. Fluroquinolones are a class of antibiotics sometimes used to treat acute bacterial sinusitis, bronchitis, and urinary tract infections; though due to the safety concerns, the current label states they should only be prescribed when patients have no alternative treatment options. RWD studies significantly contributed to the body of evidence that supported the FDA's decision to add boxed warnings regarding tendon rupture associated with fluoroquinolone use in 2008, for worsening symptoms of myasthenia gravis in 2010, and for irreversible peripheral neuropathy in 2013. RWD was also displayed at an FDA advisory committee meeting in 2015 demonstrating fluoroquinolones were widely used despite current restrictive labeling.[13] The meeting evaluated results from multiple real-world studies assessing safety issues spanning peripheral neuropathy,[14] retinal detachment,[15] tendon rupture,[16] cardiac arrhythmia,[17] and aortic aneurysm.[18] These RWD studies complemented data from clinical trials and spontaneous adverse event reports, providing a more comprehensive picture of the safety profile of fluoroquinolones in real-world use resulting in multiple regulatory actions aimed to evaluate the risks and benefits of these antibiotics.

Case study (safety labelling): Methotrexate

Methotrexate is used to treat a variety of conditions including certain forms of lymphoma, breast cancer, as well as certain autoimmune diseases, such as rheumatoid arthritis and refractory psoriasis. Though effective, the drug is associated with high toxicity and severe side effects, including death, especially when taken too frequently. In 2019, the FDA leveraged its Sentinel system to investigate dosing errors among patients with rheumatoid arthritis with new use of methotrexate. The study estimated that 0.4% of patients experienced an overdose requiring rescue therapy.[19] These findings, combined with adverse event reports, prompted the FDA to mandate changes to methotrexate's prescribing information. The required modifications included a new warning about dosing error risks, clarification of the dosing schedule for non-oncologic uses, and the development of patient medication guides.

Case study (label expansion): Ibrance

In 2019, the FDA set another significant precedent in the use of RWE by approving a supplemental New Drug Application (sNDA) for Ibrance (palbociclib) in male breast cancer patients. This approval was based on information from clinical trials supplemented by RWE derived from electronic health records in the Flatiron Health database. The RWE study demonstrated that Ibrance's safety and effectiveness profile in male patients was consistent with that observed in female patients, supporting the drug's use case for treating male patients with breast cancer.[20] The RWD-based study was able to provide evidence for expanding Ibrance's use more rapidly than would have been possible through traditional clinical trials, accelerating patient access to this treatment. The FDA's decision showcases regulatory flexibility in considering alternative evidence sources and underscores the value of large-scale EHR databases in generating regulatory-grade evidence, particularly for rare conditions or underrepresented patient groups.

Case study (label expansion): Prograf

In July 2021, the FDA made a landmark decision by approving Prograf (tacrolimus) for preventing organ rejection in lung transplant patients, based solely on RWE. This groundbreaking approval utilised data from the U.S. Scientific Registry of Transplant Recipients, comparing Prograf-based immunosuppression to cyclosporine-based regimens in adult and paediatric patients who received lung transplants.[21] The study focused on patient and graft survival at 1-year post-transplantation. Results demonstrated improved outcomes with Prograf, including higher 1-year survival rates, while maintaining a safety profile consistent with its known effects in patients who received other organ transplant. This decision not only expanded treatment options for lung transplant recipients but also set a new precedent for the use of high-quality registry data in regulatory approvals. Thus, by leveraging RWE from a large patient registry, the FDA validated the potential of RWD to inform regulatory decisions.

Conclusion

The FDA has progressively integrated RWE into its regulatory framework, evolving from its initial use in post-marketing safety monitoring to a more comprehensive approach. This evolution

Image: Freepik

was marked by key legislative milestones such as the Prescription Drug User Fee Act III in 2002 and the 21st Century Cures Act in 2016. The FDA's RWE framework emphasises transparency, data quality, and appropriate methodologies for analysing and interpreting RWD. To support this framework, the FDA has issued numerous guidance documents addressing various aspects of RWE use in regulatory decision-making. These documents provide stakeholders with clear expectations and methodological considerations for using RWE in regulatory submissions. The FDA's approach recognises the potential of RWE to inform regulatory decisions throughout a product's lifecycle, from approval of new indications to post-approval studies. The establishment of the CCRI in 2024 further demonstrates the FDA's commitment to advancing RWE use in drug development and regulatory processes. This evolving stance reflects the FDA's adaptation to a changing healthcare landscape, where insights from routine clinical practice can significantly inform regulatory decision-making.

Looking ahead, the increasing use of RWE is expected to lead to more efficient drug development, faster regulatory approvals, and improved patient outcomes by bridging the gap between clinical trials and real-world practice. As technology and data analytics continue to advance, RWE is set to become increasingly central in driving the future of healthcare and medical innovation. The integration of artificial intelligence with RWE will revolutionise our ability to identify patterns, predict outcomes, and personalise treatments. Artificial intelligence algorithms can sift through vast amounts of RWD to generate hypotheses, design more targeted clinical trials, and may even predict potential safety issues before they emerge in clinical practice. RWE will be crucial in understanding long-term efficacy and safety profiles of upcoming cell and gene therapies, given the novelty, complexity, and potentially curative nature of these treatments. As these innovative therapies often target rare diseases or specific genetic profiles, RWE can provide valuable insights into their real-world performance across diverse patient populations and healthcare settings. Moreover, RWE will be instrumental in addressing the unique challenges posed by these therapies, such as durability of response and potential long-term side effects, which may not be fully captured in traditional clinical trials. By leveraging RWE, regulators and healthcare providers are better equipped to make informed decisions regarding the utilisation, safety, effectiveness, and value of cutting-edge treatments, ultimately accelerating the path from scientific breakthroughs to patient benefit.

Acknowledgements

We would like to thank our colleagues within Medical Writing Services as well as other associated functions at Parexel who provided their insights and performed critical review of this article. Specifically, we would like to thank (listed in alphabetical order according to the first name): Carmen deKeijzer, Preeti Dhillon, Shruti Diggavi, Simin Faria, and Tina Tsai.

Disclaimers

The opinions expressed in this article are the authors' own and not necessarily shared by their employer or EMWA.

Disclosures and conflicts of interest

The authors declare no conflicts of interest.

References

1. Prescription Drug User Fee Amendments of 2002 (PDUFA III). Public Law 107-188, 116 Stat. 687 (2002).
2. FDA. FDA's Sentinel Initiative. FDA; 2023 [cited 2023 May 24]. Available from: https://www.fda.gov/safety/fdas-sentinel-initiative.
3. Kakkis ED, O'Donovan M, Cox G, et al. Recommendations for the development of rare disease drugs using the accelerated approval pathway and for qualifying biomarkers as primary endpoints. Orphanet J Rare Dis. 2015;10:16. doi:10.1186/s13023-014-0195-4
4. Sherman RE, Anderson SA, Dal Pan GJ, et al. Real-world evidence – What is it and what can it tell us? N Engl J Med. 2016;375(23):2293-7. doi:10.1056/NEJMsb1609216
5. FDA. Framework for FDA's real-world evidence program. Silver Spring (MD): FDA; 2018 Dec [cited 2023 May 24]. Available from: https://www.fda.gov/media/120060/download.
6. FDA. Real-world evidence. Silver Spring (MD): FDA; 2021 [updated 2023 May 11; cited 2023 May 24]. Available from: https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence.
7. FDA. Use of real-world evidence to support regulatory decision-making for medical devices]. Silver Spring (MD): FDA; [cited 2023 May 23]. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/use-real-world-evidence-support-regulatory-decision-making-medical-devices.

8. FDA. Submitting documents using real-world data and real-world evidence to FDA for drugs and biological products. Silver Spring (MD): FDA; 2019 [cited 2023 May 30]. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/submitting-documents-using-real-world-data-and-real-world-evidence-fda-drug-and-biological-products.
9. FDA. Real-world data: Assessing electronic health records and medical claims data to support regulatory decision-making for drug and biological products. Silver Spring (MD): FDA; 2021 [cited 2023 May 30]. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/real-world-data-assessing-electronic-health-records-and-medical-claims-data-support-regulatory.
10. FDA. Data standards for drug and biological product submissions containing real-world data. Silver Spring (MD): FDA; 2021 [cited 2023 May 30]. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/data-standards-drug-and-biological-product-submissions-containing-real-world-data.
11. FDA. Considerations for the design and conduct of externally controlled trials for drug and biological products. Silver Spring (MD): FDA; 2023 [cited 2023 May 30]. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/considerations-design-and-conduct-externally-controlled-trials-drug-and-biological-products.
12. FDA. CDER establishes new Center for Real-World Evidence Innovation. Silver Spring (MD): FDA; 2023 Oct 2 [cited 2023 Nov 14]. Available from: https://www.fda.gov/drugs/drug-safety-and-availability/cder-establishes-new-center-real-world-evidence-innovation
13. Nambiar S. FDA Introductory Remarks. Joint Meeting of the Antimicrobial Drugs Advisory Committee and the Drug Safety and Risk Management Advisory Committee; 2015 Nov 5; Silver Spring, MD. Silver Spring (MD): US Food and Drug Administration; 2015.
14. Etminan M, Brophy JM, Samii A. Oral fluoroquinolone use and risk of peripheral neuropathy: a pharmacoepidemiologic study. Neurology. 2014;83(14):1261-3. doi:10.1212/WNL.0000000000000846
15. Chui CS, Wong IC, Wong LY, et al. Association between oral fluoroquinolone use and the development of retinal detachment: a systematic review and meta-analysis of observational studies. J Antimicrob Chemother.2015;70(4):971-8. doi:10.1093/jac/dku507
16. van der Linden PD, Sturkenboom MC, Herings RM, et al. Increased risk of achilles tendon rupture with quinolone antibacterial use, especially in elderly patients taking oral corticosteroids. Arch Intern Med. 2003;163(15):1801-7. doi:10.1001/archinte.163.15.1801
17. Chou HW, Wang JL, Chang CH, et al. Risks of cardiac arrhythmia and mortality among patients using new-generation macrolides, fluoroquinolones, and β-lactam/β-lactamase inhibitors: a Taiwanese nationwide study. Clin Infect Dis. 2015;60(4):566-77. doi:10.1093/cid/ciu914
18. Lee CC, Lee MT, Chen YS, et al. Risk of aortic dissection and aortic aneurysm in patients taking oral fluoroquinolone. JAMA Intern Med. 2015;175(11):1839-47. doi:10.1001/jamainternmed.2015.5389
19. Herrinton LJ, Woodworth TS, Eworuke E, et al. Development of an algorithm to detect methotrexate wrong frequency error using computerized health care data. Pharmacoepidemiol Drug Saf. 2019;28(10):1361-8.
20. Pfizer. U.S. FDA Approves IBRANCE® (palbociclib) for the Treatment of Men with HR+, HER2- Metastatic Breast Cancer. Updated 2019 Apr 4 [cited 2025 Jul 27]. Available from: https:/www.pfizer.com/news/press-release/press-release-detail/u_s_fda_approves_ibrance_palbociclib_for_the_treatment_of_men_with_hr_her2_metastatic_breast_cancer
21. FDA. FDA approves new use of transplant drug based on real-world evidence. FDA; 2021 Jul 16 [cited 2023 May 24]. Available from: https://www.fda.gov/drugs/news-events-human-drugs/fda-approves-new-use-transplant-drug-based-real-world-evidence.

Author Information

Shadaab Ayub, M.Pharm (Pharmacology), is a Principal Medical Writer and a subject matter expert in safety documents within Medical Writing Services at Parexel International and is based in Hyderabad, India. His expertise extends to various aspects of regulatory medical writing, including the preparation and review of critical documents. He is passionate about real-world evidence and technological innovations, exploring their potential to enhance and their application in medical writing.

Scott Swain, PhD, MPH, is an Executive Director of Epidemiology at Parexel International, based in Kansas, USA. Dr Swain's career spans nearly two decades across academic, consulting, FDA, and CRO settings, positioning him as an experienced strategic advisor in real-world evidence strategies to support medical product development and FDA approval. Throughout his career, Dr Swain has led and supported numerous real-world evidence studies across a wide range of therapeutic areas, with particular expertise in oncology, haematology, and Type 1 diabetes research, contributing significantly to the advancement of medical knowledge and patient care in these fields.

EMWA NEEDS YOU

EMWA is a member-run organisation

When you volunteer to assist EMWA in any capacity, you are furthering the development of our association.

You can choose how you want to get involved: in a very limited way or as part of a larger project. The choice is yours, and everyone shares the benefits.

EMWA members can volunteer in the following areas:

Conference
- Planning Committee
- Advertising

Finance

Journal
- Submitting articles

Website
- Contributions
- Web team

Freelance Business Group

Social Media Team

Training
- Leading workshops
- Professional development
- Webinar contributions
- Webinar team

Special Interest Groups
- Entrepreneurship
- Communicating with the Public
- Medical Communications
- Medical Devices
- Pharmacovigilance
- Regulatory Writing
- Sustainability
- Veterinary Medical Writing

Ambassador Programme

Getting Into Medical Writing Group

Executive Committee
- President
- President-Elect
- Journal Editor
- Public Relations Chair
- Conference Chair
- Honorary Secretary
- Education Officer
- Treasurer
- EMWA Web Manager

WHY VOLUNTEER?
- Help promote the role of medical writers and strengthen our association
- Help to raise standards in our field
- Increase your visibility and communication opportunities within the medical writing community
- Add some prestige to your CV
- Improve your knowledge of medical writing and related topics

TO FIND OUT MORE

If you are a member of EMWA and eager to support ongoing initiatives, please contact info@emwa.org.

Expanding the safety horizon: How real-world evidence shapes drug safety

Andrew Balkin[1]
Maria Kołtowska-Häggström[2]
[1] Ambermed Medical Writing and Consulting, London, United Kingdom
[2] Proper Medical Writing, Warsaw, Poland

doi: 10.56012/qsvn4434

Correspondence to:
Andrew Balkin
andrew@ambermed.org

Abstract

Real-world evidence (RWE), generated from real-world data (RWD), is pivotal in evaluating the safety and effectiveness of medical treatments beyond the controlled settings of clinical trials. Unlike randomised controlled trials (RCTs), which often involve homogeneous patient populations and limited follow-up periods, RWD utilises diverse data sources, such as electronic health records, insurance claims, and patient registries, to assess safety and treatment outcomes in the general population. Safety data derived from RWD are critical for post-market surveillance, long-term safety monitoring, and the identification of rare or delayed adverse events. Furthermore, RWE provides insights into drug interactions and treatment effectiveness across varied demographic groups, including those underrepresented in clinical trials. Despite challenges related to data quality, confounding variables, and causal inference, RWE plays a crucial role in ensuring continuous safety monitoring and informing regulatory decisions post-approval.

In the continually advancing field of medicine, ensuring the safety and effectiveness of treatments remains a (or more likely *the*) primary concern. Clinical trials have long been regarded as the gold standard for assessing the safety and efficacy of novel treatments. These controlled research trials offer valuable insights into a drug's performance under precisely defined conditions. However, despite their critical role, clinical trials often fail to capture the full range of risks and benefits encountered in real-world practice, and are often restricted to a select patient population under controlled conditions. Real-world data (RWD), in contrast, derives information from routine clinical practice, offering a more extensive and representative perspective on a drug's safety and performance. RWD has become increasingly vital for monitoring the safety of medical interventions including pharmaceutical treatments once they are approved and used in broader populations.[1–3]

Clinical trials often fail to capture the full range of risks and benefits encountered in real-world practice.

What Is real-world evidence?

Real-world evidence (RWE) is based on RWD, which includes patient health information and healthcare delivery data routinely collected from a variety of sources such as electronic health records (EHRs), insurance claims data, patient registries, mobile health applications, and wearable devices.[2] RWE utilises these data to assess how a treatment performs in the patient population at large, as opposed to the often highly selective cohort involved in randomised controlled trials (RCTs). RWE also provides epidemiological information about adverse events, atypical treatment reactions and constitutes the basis for safety signalling.

Typically, RWD are observational, meaning they do not involve experimental interventions. Instead, they analyse the health outcomes of patients treated with a particular drug or intervention under standard clinical conditions. This methodology helps uncover valuable information about the treatment, including long-term effects, benefits, and potential risks.[3]

The role of safety data in real-world evidence

The role of safety data within RWE is particularly significant, as it provides a more diverse and comprehensive dataset compared with traditional clinical trials. The importance of safety data in RWE is evident in several key areas:

Post-market surveillance

Once a drug or device is approved by regulatory authorities, continuous monitoring of its safety and effectiveness in real-world conditions is necessary. RWE, drawing on data from a broader and more varied population, can identify adverse events or rare side effects that may not have been detected during the clinical trial phase.[4] (Figure 1) Additionally, real-world safety data can uncover potential drug–drug interactions

Image: Freepik

that were not identified in controlled clinical trial settings.

Post-market surveillance is crucial once a drug is approved and enters the market. RWD provides an effective mechanism for ongoing surveillance, allowing healthcare providers, patients, and regulatory authorities, to track adverse events and safety signals as they arise.[5] Without continuous monitoring, safety concerns may not be detected until they affect a large number of patients. In the past, several drugs have been withdrawn from the market after post-marketing surveillance revealed previously unrecognised risks. RWE serves as an ongoing safeguard, allowing regulatory bodies to take timely action when new safety issues emerge.[6]

Currently, EU drug regulations can require the collection of RWD as a condition for marketing authorisation. Such safety data collection is carried out through non-interventional post-authorisation safety studies (PASS). Even when not mandatory, PASS may be recommended to pharmaceutical companies, and they can then decide whether to conduct them. The design of PASS should be carefully considered and discussed with regulatory authorities. Often, these studies not only collect general safety information but also focus on specific abnormalities or suspected adverse drug reactions identified in RCTs, e.g. liver function abnormalities, QT prolongation, or tumour growth.

> Post-market surveillance is crucial once a drug is approved and enters the market.

It is worth noting that the EMA maintains the HMA–EMA Catalogue of RWD (formerly known as the EU PAS Register), which is a unique online source of RWD and RWE. As of May 2025, the catalogue includes 246 data sources and 3,067 studies.

Long-term safety monitoring

Clinical trials generally track patients for a limited period, often ranging from a few months to a few years. However, the full spectrum of a medication's long-term effects may not become apparent until later, sometimes much later. By following patients over extended periods, RWD can identify chronic side effects or benefits that only manifest with prolonged use. Longitudinal data particularly critical for drugs intended for long-term use, such as those used to treat chronic conditions like diabetes or cardiovascular diseases.[7] PASS, as discussed in the previous section, can serve as valuable sources for publications that provide insights into the long-term safety of new medicines. One example of such a study is ACROSTUDY, a global non-interventional safety surveillance study examining the long-term treatment of acromegaly with pegvisomant, a growth hormone (GH) receptor antagonist. Based on clinical trial data, concerns were raised about potential liver function abnormalities and pituitary tumour growth. As a result, marketing authorisation was granted on the condition that PASS be conducted, which subsequently generated reassuring safety data.[8]

Diverse patient populations

Clinical trials often impose strict inclusion and exclusion criteria, resulting in a homogeneous trial population. Consequently, the findings from clinical trials may not be fully representative of how a drug performs across different demographic groups, including patients of various ages, ethnicities, or those with multiple underlying health conditions. RWE trials can include diverse populations, thereby providing a more accurate depiction of how a drug affects different segments of the population. This is especially important when assessing safety for vulnerable groups, including the elderly, pregnant women, and individuals with comorbidities.[9] An example of such data is work performed by Vila et al who studied the treatment outcomes and safety of GH replacement during pregnancy in women with GH deficiency (GHD). The Pfizer International Metabolic Database (KIMS) collected RWD in adult patients with hypopituitarism, and 201 pregnancies were identified. Based on these data, the authors concluded that there was no relationship between pregnancy and GH replacement.[10]

Another example is a mortality study conducted using RWD from the same registry. The question of whether GH replacement therapy improves life expectancy in patients with GHD had remained unanswered for a long time. Gaillard et al. published RWE on mortality in patients with various underlying causes of GHD. Their findings showed that patients with hypopituitarism due to craniopharyngioma or aggressive tumours continued to exhibit increased mortality rates, while those with other underlying causes had life expectancies comparable to the general population.[11]

A different approach to studying mortality in a specific patient group was taken by Shankar et al., who investigated patients with drug-resistant epilepsy (DRE) using data from the Clinical Practice Research Datalink (CPRD).[12] CPRD is a real-world research database that collects de-identified primary care data from a network of general practices across the UK. Since DRE is not coded as a distinct diagnosis in the database, the authors defined specific criteria to identify this patient cohort. Based on this methodology, they were able to assess the prevalence of co-morbidities, as well as all-cause and epilepsy-related mortality, between January 1, 2011 and March 31, 2021. Their findings showed that the mortality rate in patients with DRE was approximately four times higher than that of the general population in the UK.

Improved understanding of drug interactions

Clinical trials typically involve a narrow patient population that meets specific inclusion criteria, leading to a homogeneous sample. As a result, the findings from clinical trials may not be fully generalisable to the broader, more diverse

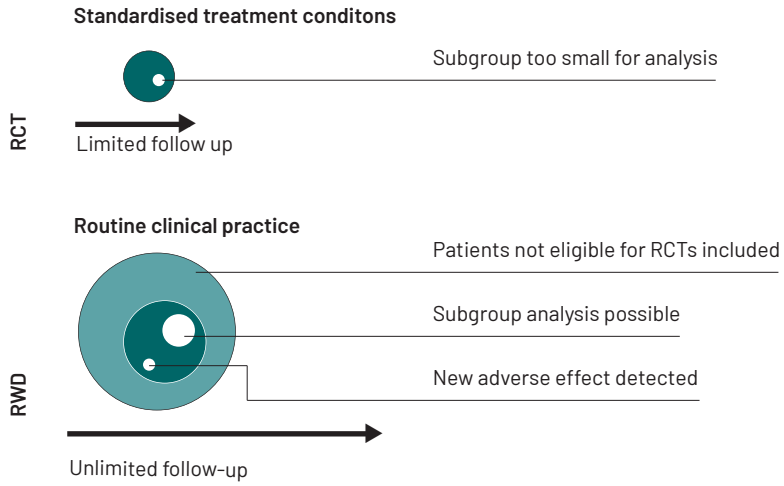


Figure 1. Sources of safety information

Abbreviations: RCT, randomised clinicial trial; RWD, real-world data.

Difference between RCT and RWD, courtesy of the late Dr Berhard Saller.

population. For example, a medication that shows efficacy in young, healthy participants in clinical trials may exhibit a different safety profile when used by older individuals or patients with multiple comorbidities.[13] RWE can reveal how demographic and health factors influence the safety and effectiveness of treatments in real-world populations.

Detection of rare and long-term adverse events

Certain adverse events are infrequent and may not be detectable in the relatively small sample sizes of clinical trials. (Figure 1) By utilising large, real-world datasets, RWE trials can identify rare but potentially serious side effects that may not have been apparent in the clinical trial phase. For instance, some adverse reactions such as specific cancers or severe allergic responses may only occur in a small subset of patients but could have significant health implications if not identified and addressed in a timely manner.[14] For example, meta-analyses found that insulin therapy seemingly has an association with increased overall cancer risk, and has significant associations with colorectal and pancreatic cancers.[15–17]

Although clinical trials are effective at detecting common side effects, they often lack the statistical power to identify rare or infrequent adverse events. RWD, which involve larger and more heterogeneous patient populations, are better equipped to detect these rare occurrences. For example, severe allergic reactions may affect only a small subset of patients in clinical trials, but these adverse events may become more apparent as the drug is used by a larger and more diverse population. Furthermore, certain safety issues – such as organ toxicity or cardiovascular complications – may not emerge until years after the initiation of treatment. RWD provide the longitudinal data necessary to monitor these long-term effects, thus offering a more complete picture of a drug's safety profile.[18]

Regulatory bodies and safety data

Regulatory agencies such as the EMA and the FDA acknowledge the growing importance of RWE in post-market surveillance. Both agencies are increasingly relying on RWE to monitor the safety of drugs and medical devices once they are available to the public. For example, the FDA has established guidelines for using RWE to support the approval of new indications for drugs and to assess ongoing safety.[19,20]

> Regulatory agencies acknowledge the growing importance of RWE in post-market surveillance.

The collection of safety data through RWE also enables reglatory authorities to act swiftly if new safety concerns arise. Should a concerning trend, such as a rise in adverse event reports, be identified, regulatory agencies can take timely action, including issuing warnings, modifying labelling, or even withdrawing a product from the market.[21]

Finally, speaking of long-term registries, it is impossible to omit the pioneering RWD registry run between 1987 and 2012; this was initiated at the request of regulatory authorities as a post-marketing surveillance study to follow 500 patients treated with GH (Genotropin®) for 5 years. It has evolved to be one of if not the largest and longest-running pharmaco-epidemiological study with four primary objectives:

1. to evaluate the long-term safety of GH and GH treatment outcomes in subjects who were treated with Genotropin®;
2. to determine relationships between clinical status, dosage schedule, and response to Genotropin® treatment;
3. to develop clinical tools for individualised GH treatment of children;
4. to contribute to the knowledge of growth and growth disorders.

It was a unique source of knowledge that was shared within the public domain and yielded 129 publications, cited in PubMed.[22]

Conclusion

RWD and corresponding RWE serve as a critical supplement to traditional clinical trials, providing valuable insights into the safety and effectiveness of treatments within diverse, real-world populations. By continuously monitoring safety data and identifying potential risks, RWE ensures that medical treatments continue to benefit patients well beyond the initial approval phase. As healthcare systems globally integrate RWD, the role of safety monitoring and post-market surveillance will continue to expand, ultimately ensuring that patients receive the safest and most effective treatments available.[23]

The examples presented in this article highlight the undeniable value of RWE in enhancing our understanding of complex medical conditions, guiding optimal therapeutic approaches, and improving everyday clinical practice.

Disclosures and conflicts of interest

The authors declare no conflicts of interest.

References

1. Murphy SA, D'Ambrosio C. Real-world evidence in clinical research: methodologies and applications. JAMA. 2021;325(7):663-73.
2. Kern DM, McLaughlin M. Use of real-world evidence in post-marketing surveillance of pharmaceuticals. N Engl J Med. 2019;380(12):1114-25.
3. Sherman RE, Anderson SA, Dal Pan G. Real-world evidence – what is it and what can it tell us? N Engl J Med. 2016;375(23):2263-7. doi:10.1056/NEJMsb1609216
4. FDA. Framework for FDA's real-world evidence program. 2018. [Accessed 2025 Mar 02]. Available from: https://www.fda.gov/media/120060/download
5. EMA. Use of real-world data and evidence in the evaluation of medicines. 2020. [Accessed 2025 Mar 02]. Available at: https://www.ema.europa.eu/en/about-us/how-we-work/data-regulation-big-data-other-sources/real-world-evidence
6. Langan SM, Smeeth L. Real-world evidence in healthcare: the role of big data and machine learning. Lancet Digit Health. 2019;1(3):e140-50.
7. Pereira M, Ramlall R. Real-world evidence in pharmacovigilance: a critical overview. Drug Saf. 2020;43(6):591-602.
8. van der Lely AJ, Gutson RK, Trainer PJ, et al. Long-term treatment of acromegaly with pegvisomant, a growth hormone receptor antagonist. Lancet. 2001;24;358(9295):1754-9. doi:10.1016/S0140-6736(01)06844-1
9. Herrett E, Thomas SL. The role of real-world data in the evaluation of drug safety. BMJ. 2017;358:j3142.
10. Vila G, Akerblad AC, Mattson AF, et al. Pregnancy outcomes in women with growth hormone deficiency. Fertil Steril. 2015;104(5):1210-7.e1.
11. Gaillard RC, Mattson AF, Akerblad AC, et al. Overall and cause-specific mortality in GH-deficient adults on GH replacement. Eur J Endocrinol. 2012;166(6):1069-77. doi:10.1530/EJE-11-1028
12. Shankar R, Marston XL, Danielson V, et al. Real-world evidence of epidemiology, patient characteristics, and mortality in people with drug-resistant epilepsy in the United Kingdom, 2011-2021. J Neurol. 2024;271(5):2473-83. doi:10.1007/s00415-023-12165-4
13. Hernán MA, Robins JM. Causal inference: what if. Chapman & Hall/CRC;2016.
14. D'Agostino RB, Rubin DB. Observational studies and their role in evidence generation. Lancet. 2021;398(10296):61-70.
15. Janghorbani M, Dehghani M, Salehi-Marzijarani M. Systematic review and meta-analysis of insulin therapy and risk of cancer. Horm Cancer. 2012;3:137-46. doi:10.1007/s12672-012-0112-z
16. Li D, Tang H, Hassan MM, et al. Diabetes and risk of pancreatic cancer: a pooled analysis of three large case–control studies. Cancer Causes Control [BMG-4]. 2011;22:189-197. doi:10.1007/s10552-010-9686-3
17. Deng L, Gui Z, Zhao L, et al. Diabetes mellitus and the incidence of colorectal cancer: an updated systematic review and meta-analysis [BMG5]. Dig Dis Sci. 2012;57:1576-85. doi:10.1007/s10620-012-2055-1
18. Buchan IE, de Lusignan S. The power of real-world evidence in medicine: applications and limitations. Health Policy and Technology [BMG-6] 2020;9(1):1-10.
19. Zhang W, Li M. Real-world evidence for drug-drug interactions: a new paradigm. J Clin Pharmacol. 2021;61(2):217-25.
20. EMA.Heads of Medicines Agency. Real-world evidence framework to support EU regulatory decision-making: Report on the experience gained with regulator-led studies from September 2021 to February 2023. 2023. [Accessed 2025 Mar 02]. Available at: https://www.ema.europa.eu/en/documents/report/real-world-evidence-framework-support-eu-regulatory-decision-making-report-experience-gained-regulator-led-studies-september-2021-february-2023_en.pdf
21. Wang X, Ma Y. The role of real-world data in drug approval and post-marketing surveillance. J Clin Trials. 2019;6(4):213-20.
22. Geffner ME, Ranke MB, Wajnrajch MP, et al. An overview of growth hormone therapy in pediatric cases documented in the Kabi International Growth Study (Pfizer International Growth Database). Ann Pediatr Endocrinol Metab. 2024;29(1):3-11. doi:10.6065/apem.2346206.103
23. Riley TR, Smith LC. Evaluating drug safety with real-world data: a multi-country approach. Int J Med Inform[BMG-7]. 2020;135:104078.

Author information

Andrew Balkin has been a medical writer for more than 20 years, and recently co-founded Ambermed Medical Writing and Consulting.

Maria Kołtowska-Häggström, MD, PhD, runs Proper Medical Writing, the first Polish medical writing agency. She has 30 years of experience in the pharmaceutical industry, primarily in clinical research, including real-world evidence (RWE).

Defining the quality of data within rare disease registries: A systematic review

Jessica Anderson, Malika Alimussina, S. Faisal Ahmed
Office for Rare Conditions, School of Medicine, Dentistry & Nursing, University of Glasgow, UK

doi: 10.56012/knxk2392

Correspondence to:
S. Faisal Ahmed
Faisal.ahmed@glasgow.ac.uk

Abstract

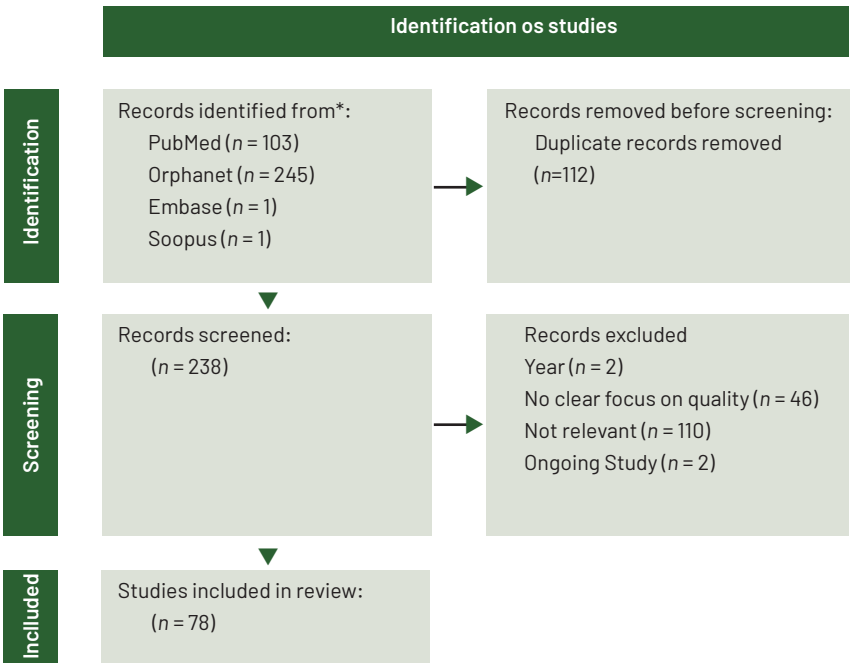
Rare diseases have a low prevalence within society, resulting in limited awareness and challenges with data availability for research. While rare disease registries offer valuable data, ensuring quality of the data is essential. This review explores key themes and influencing factors affecting data quality in rare disease registries. Studies were identified through a pre-defined search term across multiple databases and screened for recurring themes and terms. The findings indicate a growing emphasis on data quality and evolving perspectives on how it is defined and assessed through the years.

Rare diseases are often defined as conditions that affect fewer than 5 out of 10,000 members of the general population[1] and these conditions may affect up to 6%-7% of the world's population.[2] The low prevalence of these conditions often leads to limited awareness of the conditions as well as their management among both the public and healthcare professionals.[2] The lack of data affects the development of an adequate amount of evidence that can inform safety and effectiveness of drugs, diagnosis, and research in general. To address these challenges, one possible solution is the development of patient registries.

Patient registries are databases that are designed to systematically collect, store, and analyse clinical data. They can be used to track patient demographics, diagnosis, treatments, and outcomes, enabling longitudinal studies to take place on a large scale. The data that are collected in these registries often represent a setting that is beyond what is encountered in controlled clinical trials or experimental environments. So these patient registries not only address the challenges of limited and heterogenous data, but they also collect real-world data that reflects how people utilise healthcare services and respond to interventions in their everyday lives.[3,4] Real-world data provide insights into disease progression, treatment outcomes, and patient experiences, which are essential for informing healthcare policy, improving clinical care, development of new drugs and interventions, monitoring the use of these interventions and for performing comparative effectiveness research.[5] As rare disease registries start to play an increasingly pivotal role in rare disease research, the rare disease community has seen a proliferation of these registries and this has an implication on long-term sustainability of these platforms.

The critical factor that will influence the long-term sustainability of a rare disease registry will be its quality and this can be broadly divided into two categories. The first one relates to its operation systems and the second category, which is equally important, relates to the data that the registry collects.[6] This is even more important in rare diseases where the populations are very small and poor data quality may skew the results or lead to inconclusive results thus limiting the acceptability of the findings. Data quality itself may be defined in several ways including completeness, interoperability, accuracy, validity, consistency, timeliness, uniqueness and traceability.[7,8] Amongst existing registries, it is clear that the definition of registry quality may be quite variable[9] and the level of consensus that may exist for data quality is also unclear. It is important to understand the key concepts of data quality so that resources can be directed towards these to ensure long-term sustainability. Furthermore, registries with a higher level of data quality are more likely to have greater acceptability amongst health care providers. The current systematic review was, therefore, performed to explore the key concepts of data quality that are reported in contemporary rare disease registry literature.

Figure 1. PRISMA flowchart outlining the inclusion exclusion criteria used to screen the literature



Methods

A systematic review was performed to examine how data quality is defined in rare disease registries by synthesising literature from 2010 to 2025 and identifying key themes and related components that define data quality. Thematic analysis was performed to categorise recurring themes and trends that were observed within the literature. The inclusion criteria included publications that were published in English in a peer reviewed journal from 2010 onwards and had a clear focus on data quality and rare disease registries. The 15-year time period was chosen as it was felt to be a relevant period to capture a sufficient amount of literature within the field. Rare diseases were included in the criteria to ensure the relevant population was captured appropriately. Non-peer reviewed literature was excluded to ensure the reliability of the literature for this analysis. The systematic review was conducted and reported in accordance with the method outlined in the Cochrane Handbook for Systematic Reviews[10] and Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.[11] Literature search results were uploaded to Covidence (Covidence systematic review software, Veritas Health Innovation, Melbourne, Australia). Articles were manually screened by title and abstract to determine eligibility according to the inclusion criteria above. Relevant full-text studies were collated and evaluated for eligibility for inclusion (Figure 1). The selected studies were also screened for the definitions and themes as previously described.[12] These data were then extracted from Covidence for frequency analysis of the definitions of data quality and the factors that affect data quality. Lastly, thematic analysis was performed to identify recurring themes within the literature. Following initial familiarisation with the literature within the field, key concepts and phrases were identified (Figure 2). The thematic analysis was used to identify trends in defining data quality in rare disease registries and trends in factors that may influence data quality in rare disease registries over the last 15 years. These temporal trends were arbitrarily divided into four time periods of three years each. The co-occurrence of themes was analysed using R, employing the tidyverse, igraph, and ggraph packages. Each article was assigned a unique article ID to facilitate tracking. Themes associated with each article ID were identified, and pairwise co-occurrences of themes within individual articles were computed. These co-occurrences were then aggregated across all articles to assess the frequency of theme co-occurrence throughout the data set.

Table 1. Themes and terms

Theme	Term
Completeness	Complete Completed Completeness
Selection Bias	Bias Selection bias
Validity	Validity Valid Validate
Accuracy	Accuracy Accurate
Interoperability Duplication	Interoperability Duplicate Duplication
Standardisation	Standardisation Standardised
Common Data Set Elements	Common data set elements MDS

Themes and their corresponding codes. It is important to note that for standardisation both the American and British spellings were used to screen the literature.
Abbreviation, MDS, minimum data set

Results

Frequency of definitions of data quality

A total of 78 studies were included, and within these studies 9 themes were identified: completeness, selection bias, validity, accuracy, consistency, interoperability, duplication, standardisation, and common data set elements. These 9 themes were further subdivided into terms that represented those that were used within these themes (Table 1). On the other hand, terms such as common data set elements, minimum data set (MDS) were not very frequent.

Trends in definitions of data quality

The total number of term occurrences grew steadily from 50 in 2010–2013 to 876 in 2022–2025, representing a 17.5-fold increase over the study period (Table 2). Terms related to completeness (e.g. *completeness*, *complete*, *completed*) were among the most frequently cited, with *completeness* alone appearing 224 times, followed by *complete* (166 times) and *completed* (90

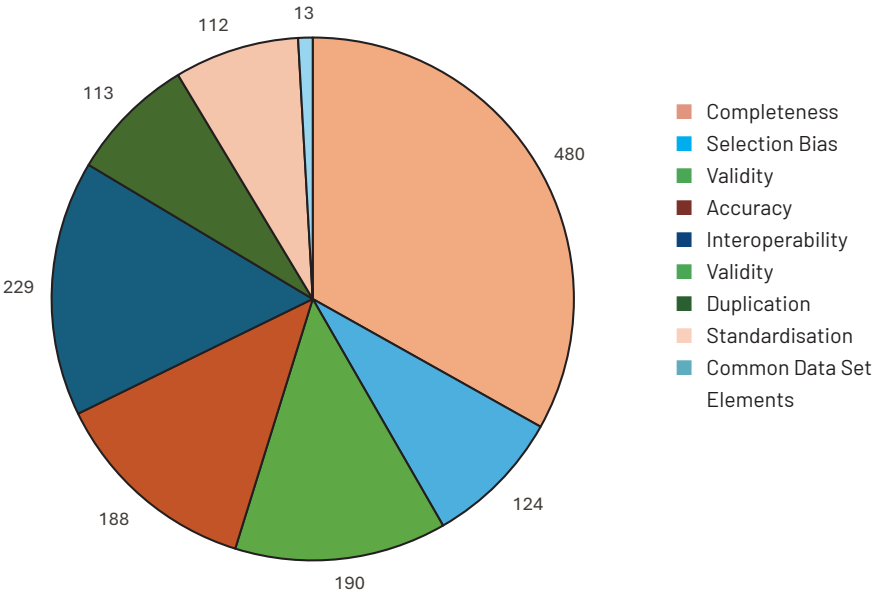


Figure 2. Theme frequency pie chart describing the frequency of terms used to screen the literature

Table 2. Temporal trends in the reporting of themes and terms, 2010–2025

Theme	Term	Temporal Group 2010–2013	2014–2017	2018–2021	2022–2025	Total for Term
Accuracy	Accuracy	0	26	30	59	115
	Accurate	0	13	24	36	73
	Total	0	39	54	95	188
Selection Bias	Bias	2	14	17	55	88
	Selection Bias	8	6	5	17	36
	Total	10	20	22	72	124
Common Data Set Elements	Common data set elements	0	0	0	0	0
	MDS	0	0	6	7	13
	Total	0	0	6	7	13
Completeness	Complete	13	20	37	96	166
	Completed	15	7	20	48	90
	Completeness	1	66	40	117	224
	Total	29	93	97	261	480
Consistency	Consistency	2	11	26	58	97
	Consistent	2	6	17	47	72
	Total	4	17	43	105	169
Duplication	Duplicate	0	21	44	14	79
	Duplication	1	7	13	13	34
	Total	1	28	57	27	113
Interoperability	Interoperability	0	5	79	145	229
Standardisation	Standardisation	1	2	12	15	30
	Standardised	2	9	17	54	82
	Total	3	11	29	69	112
Validity	Valid	0	4	12	12	28
	Validate	2	3	15	17	37
	Validity	1	38	20	66	125
	Total	3	45	47	95	190
Total Terms		**50**	**258**	**434**	**876**	**1618**

Abbreviation: MDS, minimum data set

times). Similarly, *interoperability* experienced a marked increase, from no mentions before 2014 to 145 mentions in 2022–2025, making it the most cited individual term overall (229 total mentions). Conversely, certain terms such as *common data set elements* and *MDS* were rarely mentioned or not at all, indicating either limited focus or a preference for alternative terminology. With the exception of the term *common data set elements* phrase, all other terms were present from 2018 onwards (Figure 3). From 2010 to 2013, the most frequently occurring terms were *completed* and *completeness*, marking completeness as the dominant theme in that early period. In the following period, 2014–2017, *completeness* remained the most frequent term, but it was followed closely by *validity*. A more marked shift occurred in 2018–2021, when *interoperability* and *duplicate* became the most frequently mentioned terms. By 2022–2025, the top two terms were again *interoperability* and *completeness*.

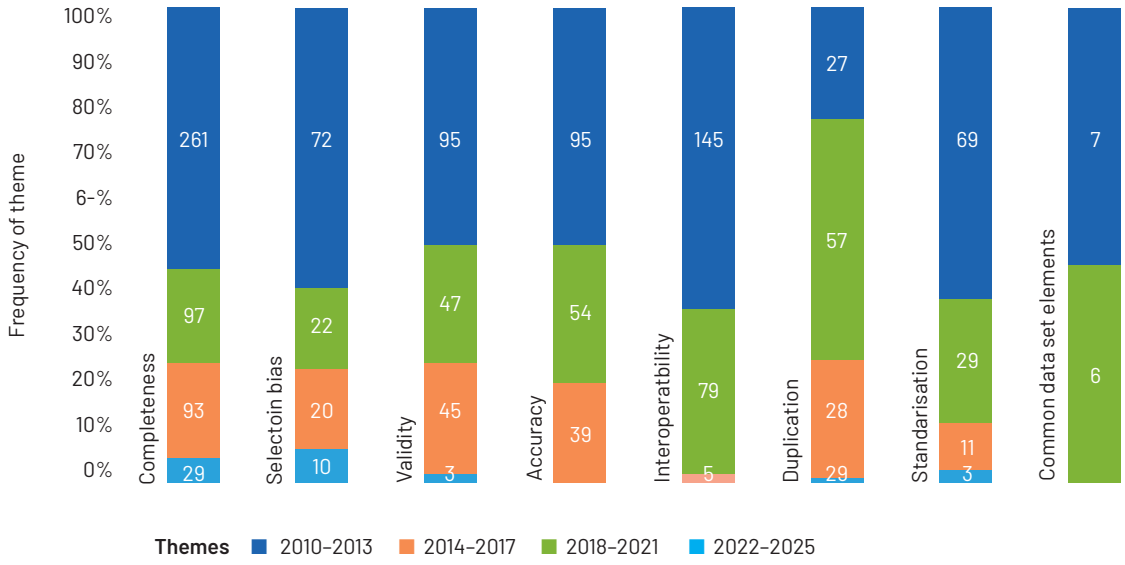


Figure 3. The frequency of themes for each of the year groups previously defined

Relationship of themes to each other

All themes co-occurred with at least one other, demonstrating that each theme had been discussed alongside others at some point in the literature (Figure 4). Whilst the theme *common data set elements* did not have a high overall frequency, it was still well interconnected. This was because the articles that had this theme also had multiple other themes occurring at the same time as well. This means that whilst the theme overall was not frequent in the literature, it was interconnected with the other themes.

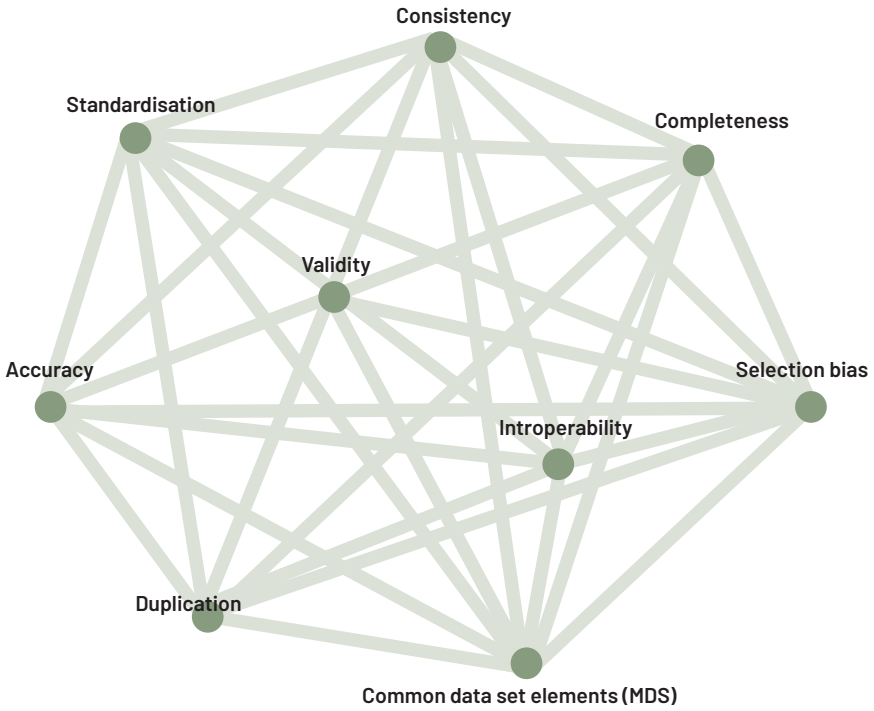


Figure 4. Theme network graph showing how interlinked each of the themes are with one another.

Discussion

This review set out to explore how data quality is defined within rare disease registries by analysing literature published between 2010 and 2025, with the aim of identifying key themes and influencing factors. Using thematic analysis framework,[12] nine recurring themes were identified across the included studies: completeness, selection bias, validity, accuracy, consistency, interoperability, duplication, standardisation, and common data set elements. Together, these themes reflect the complexity of data quality and the range of priorities currently shaping the field.

The findings show a clear progression in how data quality has been approached over time. Between 2010 and 2013, the focus tended to be on more basic aspects of quality – particularly completeness and whether data had been fully recorded – highlighting an early concern with ensuring registries captured the full picture. From 2018 onwards, however, the emphasis has shifted towards more system-level issues such as interoperability and duplication. This change points to a deeper and more technical understanding of what makes data useful, particularly when it is shared across settings or used for secondary purposes. The growing frequency of terms over time reflects an increasing interest in defining and improving data quality across both academic and clinical contexts. In addition, the overlap between themes – demonstrated through co-occurrence – suggests that these concepts are not being

Image: Freepik

considered in isolation, but as part of a broader, interrelated understanding of quality. This highlights the interconnectedness of these concepts and suggests that the definition of data quality within rare disease registries is inherently multidimensional.

Health care professionals face a variety of barriers to participating in rare disease registries. Many health care professionals are not aware of rare disease registries and even when they are aware of these registries their level of participation is limited.[13] Clinicians and associated administrative and care staff often have heavy workloads, leaving little time for data entry or patient follow-up.[14] If data were sufficiently interoperable, they could flow between different sources and the need for manual entry that may also lead to transcription errors could be minimised. However, even if this was possible, it is likely that at an institutional level, without local approval, free data flow for highly sensitive data will be challenging. Rare disease registries rarely need to collect that are real-time, and a solution for addressing the time constraints is to develop systems that can bulk download source data and subsequently upload the data at a time that is convenient. However, this still requires the need to agree on standardised data sets that can be collected universally. These data sets are referred to in different ways in the literature including common data elements,[15] core outcome sets,[16] and minimum data sets.[17] By minimising the amount of data that is collected in rare disease registries, projects such as GloBE-Reg, a global registry for novel therapies in rare bone and endocrine conditions, are aiming to improve the data quality.[17] One potential limitation of this study is the possibility that not all relevant terms such as *common data elements* or *minimum data set* frequencies were captured. This is likely due to the terms being used to search the literature not capturing the frequency of these themes accurately, potentially introducing bias.

Overall, the findings from this review highlight both an increased focus on data quality in rare disease registries over time and a shift in how quality is being conceptualised. While earlier studies primarily emphasised completeness and validity, more recent literature places greater attention on themes such as interoperability, duplication, and consistency. This shift suggests a growing and more nuanced understanding of what makes data high quality.

Disclosures and conflicts of interest

The authors are supported by GenSci, Novo Nordisk and Pfizer for the GloBE-Reg project. The authors declare no conflicts of interest.

Data availability statement

For inquiries about data and other supplemental information, please contact the corresponding author.

Refernces

1. European Commissions. Rare diseases - European Commission. 2023 [cited 2025 Jan 29]. Available from: https://health.ec.europa.eu/rare-diseases-and-european-reference-networks/rare-diseases_en
2. Nguengang Wakap S, Lambert DM, Olry A, et al. Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. Eur J Hum Genet. 2020;28(2):165–73. doi:10.1038/s41431-019-0508-0
3. Hageman IC, van Rooij IALM, de Blaauw I, et al. A systematic overview of rare disease patient registries: challenges in design, quality management, and maintenance. Orphanet J Rare Dis. 2023;18(1):106. doi:10.1186/s13023-023-02719-0
4. Shourick J, Wack M, Jannot A. Assessing rare diseases prevalence using literature quantification. Orphanet J Rare Dis. 2021;16(1):139. doi:10.1186/s13023-020-01639-7
5. Raycheva R, Kostadinov K, Mitova E, et al. Challenges in mapping European rare disease databases, relevant for ML-based screening technologies in terms of organizational, FAIR and legal principles: scoping review. Front Public Health. 2023;11: 1214766. doi:10.3389/fpubh.2023.1214766
6. Kodra Y, Weinbach J, Posada-De-La-Paz M, et al. Recommendations for improving the quality of rare disease registries. Int J Environ Res Public Health. 2018;15(8):1644. doi:10.3390/ijerph15081644
7. Wang J, Liu Y, Li P, et al. Overview of data quality : examining the dimensions, antecedents, and impacts of data quality. J Knowl Econ. 2024;15:(1): 1159–78. doi:10.1007/s13132-022-01096-6

8. Jonker CJ, de Vries ST, van den Berg HM, et al. Capturing data in rare disease registries to support regulatory decision making: a survey study among industry and other stakeholders. Drug Saf. 2021;44(8):853–61. doi:10.1007/s40264-021-01081-z
9. Ali SR, Bryce J, Kodra Y, Taruscio D, et al. The quality evaluation of rare disease registries-an assessment of the essential features of a disease registry. Int J Environ Res Public Health. 2021;18(22):11968. doi:10.3390/ijerph182211968
10. Higgins JPT, Thomas J, Chandler J, et al, eds. Cochrane Handbook for Systematic Reviews of Interventions version 6.5 (updated August 2024). doi:10.1002/9781119536604
11. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71. doi:10.1136/bmj.n71
12. Kiger ME, Varpio L. Thematic analysis of qualitative data: AMEE Guide No. 131. Med Teach. 2020;42(8):846–54. doi:10.1080/0142159x.2020.1755030
13. Ali SR, Bryce J, Cools M, et al. The current landscape of European registries for rare endocrine conditions. Eur J Endocrinol. 2019;180(1):89–98. doi:10.1530/eje-18-0861
14. Kyriakou A, Dessens A, Bryce J, et al. Current models of care for disorders of sex development - results from an international survey of specialist centres. Orphanet J Rare Dis. 2016;11:155. doi:10.1186/s13023-016-0534-8
15. Ali SR, Bryce J, Smythe C, et al. Supporting international networks through platforms for standardised data collection-the European Registries for Rare Endocrine Conditions (EuRRECa) model. Endocrine. 2021;71(3):555–60. doi:10.1007/s12020-021-02617-0
16. Flück C, Nordenström A, Ahmed SF, et al. Standardised data collection for clinical follow-up and assessment of outcomes in differences of sex development (DSD): recommendations from the COST action DSDnet. Eur J Endocrinol. 2019;181(5):545–64. doi:10.1530/eje-19-0363
17. Chen SC, Bryce J, Chen M, et al. Development of a minimum dataset for the monitoring of recombinant human growth hormone therapy in children with growth hormone deficiency: a GloBE-Reg Initiative. Horm Res Paediatr. 2024; 97(4):365–73. doi:10.1159/000533763

Author information

S. Faisal Ahmed, MD, FRCPCH, received his clinical and scientific training in Edinburgh, London, and Cambridge. He was appointed as a consultant in paediatric endocrinology at the Royal Hospital for Sick Children, Yorkhill, Glasgow, in 2000 and was appointed as the Samson Gemmell Chair of Child Health at the University of Glasgow in 2012. Since 2019, he has held an additional chair at Leiden University as Professor of Endocrine Registries. His research has focussed on improving the health of people with a wide range of rare conditions, and especially those with conditions affecting sex development or growth and skeletal development. He is the Editor-in-Chief of the journal Endocrine Connections.
ORCID: 0000-0003-0689-5549

Malika Alimussina received her clinical and academic training in paediatric endocrinology in Kazakhstan and later completed her PhD in Medicine at the University of Glasgow in 2023. Dr Alimussina is currently a Clinical Scientist at the Office for Rare Conditions, University of Glasgow, where she contributes to the development and analysis of international registries for rare endocrine disorders. Her research focuses on improving long-term outcomes in individuals with Disorders of Sex Development and other rare growth and endocrine conditions. Dr Alimussina is also a faculty member at one of the European Society for Paediatric Endocrinology training schools, where she teaches and mentors early-career clinicians and researchers.

Jessica Anderson received a BSc (Hons) in Pharmacology from the University of Aberdeen in 2023 and an MSc in Precision Medicine and Pharmacological Innovation from the University of Glasgow in 2024. She is working on a PhD focused on data quality within rare disease registries, with particular emphasis on completeness, interoperability, and consistency of core clinical variables. Her research aims to develop approaches to improve the reliability of real-world data and support translational research and clinical decision-making in rare diseases.

Lessons learnt from PETHEMA's RWD research: A clinical perspective

Alfonso J. Santiago-Marí[1],
Irene Salcedo-Marín[1],
Roberto Maldonado[1],
Carmen López-Carrero[1],
David Martínez-Cuadrón[2],
Pau Montesinos [2]

[1] Fundación PETHEMA. Madrid, Spain
[2] Haematology Department, Hospital Universitari i Politècnic La Fe València, Spain.

doi: 10.56012/taiu2814

Correspondence to:
Alfonso J. Santiago-Marí
asm@fundacionpethema.es

Abstract

PETHEMA (Programa Español de Tratamientos en Hematología, or, Spanish Program for Treatments in Haematology) a leading cooperative group in haematological research, has increasingly integrated real-world data (RWD) as a complement to traditional clinical trials. RWD research captures information from routine clinical practice, reflecting a broader patient population and enhancing external validity. PETHEMA has conducted impactful RWD studies in acute myeloid leukaemia, acute lymphoblastic leukaemia, multiple myeloma, and bone marrow failure syndromes, notably leveraging large clinical registries and molecular data. Despite challenges such as data heterogeneity, confounding factors, and regulatory constraints, RWD offers essential insights for personalised medicine. Future priorities include improving data interoperability, applying artificial intelligence, and rethinking legal frameworks to balance data protection with scientific progress. PETHEMA's experience highlights the transformative role of RWD in supporting more informed and representative clinical decision-making in real-life haematology settings.

Speaking of PETHEMA

PETHEMA (Programa Español de Tratamientos en Hematología, or, Spanish Program for Treatments in Haematology) is a cooperative research group in medicine comprised of virtually all Spanish haematologists, along with many from Portugal and other Latin American countries. Established in 1972, PETHEMA aimed to generate well-structured protocols to guide the clinical practice of haematology professionals in the field of malignant diseases. Over time, it has evolved into a cohesive and powerful group at the forefront of numerous medical research studies. These include a wide range of clinical trials, but also observational and epidemiological studies, registries, and significant basic translational research related to its focus diseases. PETHEMA operates administratively and legally through its private foundation in Madrid. This foundation currently manages over 80 diverse projects, making PETHEMA the leading non-commercial promoter of clinical studies among Spanish medical societies in Europe. It is surpassed only by three major Spanish public health network hospitals (Table 1).[1] PETHEMA primarily focuses its research activity on multiple myeloma (MM), acute myeloblastic leukaemia (AML), acute lymphoblastic leukaemia (ALL), and chronic lymphocytic leukaemia (CLL), among others.

Real world data (RWD), breaking down the boundaries of cancer research

The first cooperative research among haematologists in Spain began in the early 1970s. The primary tool for this collaboration was the care protocol. This consensus-based document outlined the key guidelines for disease diagnosis and treatment and instructed all participating clinicians to collect a set of essential data on their patients with a specific disease. This data was then sent to a national coordinator, who performed an aggregate analysis of the parameters. This analysis allowed the coordinator to formulate scientific conclusions regarding the patients' response to the pharmacological treatments used. This "virtual circle" process is depicted in the flowchart in Figure 1. This type of research was still very rudimentary, lacking today's standards (no requirement for prior authorisation, no informed consent, no perception of the need for external monitoring), but it had one great advantage: it focused on the universality of patients belonging to a given clinical profile, without incurring any special selection of patients and directly studying real daily clinical practice. In this way, one could say that this type of research was something like a prototype of modern research with real-world data (RWD), where doctors, who had agreed on common diagnostic-therapeutic procedures,

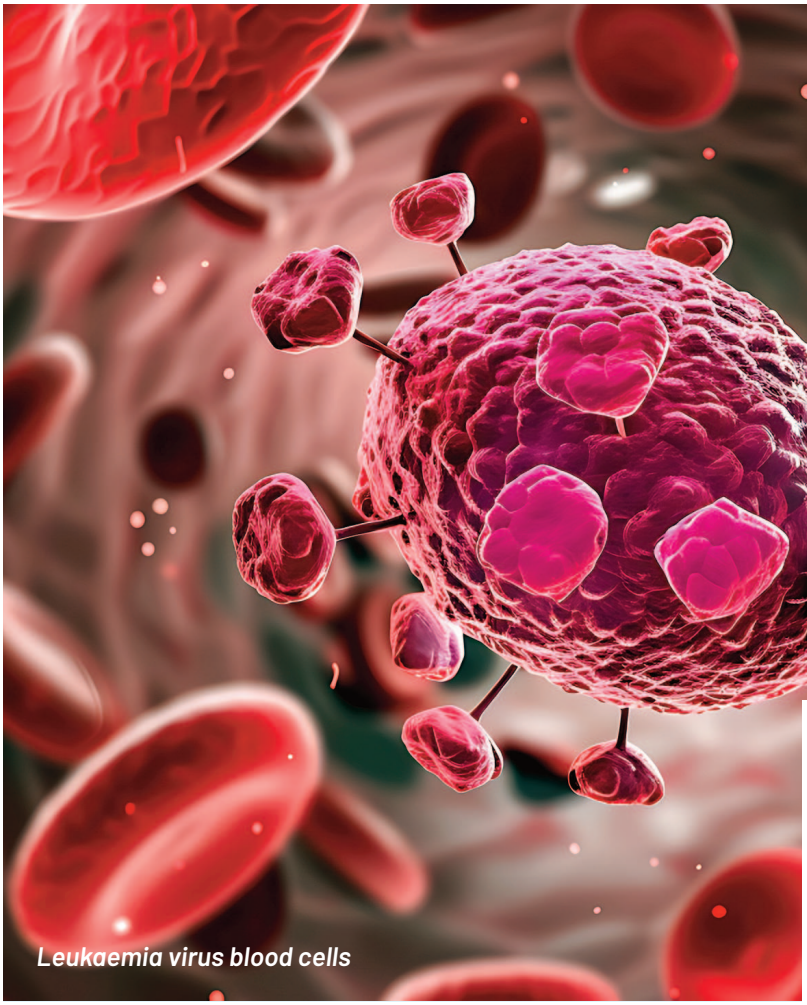
This analysis allowed the coordinator to formulate scientific conclusions regarding the patients' response to the pharmacological treatments used.

Table 1. PETHEMA (as per Oxford Index)

First Spanish medical society, non-commercial promoter of clinical studies in Medicine in European hospitals of the public health network.

Order	Sponsor	Number of studies
78	Fundació Clínic per a la Reçerca Biomèdica (Barcelona)	84 Clinical Trials
88	Vall d'Hebron University Hospital (Barcelona)	73 Clinical Trials
102	Hospital de la Santa Creu i Sant Pau (Barcelona)	67 Clinical Trials
143	PETHEMA Foundation (Madrid)	50 Clinical TrialsEU

Trials Tracker [Internet] [accessed 2025 Jun 2]. Available from: https://eu.trialstracker.net/?all



Leukaemia virus blood cells

Image: Freepik

established in parallel a real-time monitoring of their results, allowing them to agree on the continuous improvement of these same procedures.

With the continued development of research activity and the necessary legal and ethical regulation of clinical research, alongside major advancements in new drugs, clinical trials have become particularly important. In these, optimal experimental conditions are planned, and patients are selected and allocated according to rigid homogeneity criteria within a specific clinical study profile to test the efficacy and safety of new therapies. Clinical trials have been, and will continue to be, an indispensable tool for understanding the fundamental "behaviour" of new treatments in people, with a notable level of internal validity. However, in many cases, they are not sufficient to provide sufficient external validity to extrapolate their conclusions point by point to the entire population.[2]

It has therefore become necessary to go back to the origins and to recover a type of research that allowed us to obtain information from the real environment as a whole and from everyday healthcare practice with which to capture all the truthful information on the treatments used in all types of patients, i.e., RWD research. According to the agreed definition of RWD, this type of investigation includes all observational and registry studies, patient self-reported data (surveys, quality of life questionnaires, social media testimonials, among others), administrative databases, and electronic registry data (including laboratory data, digital medical records, and patient complaints, among others).[3,4]

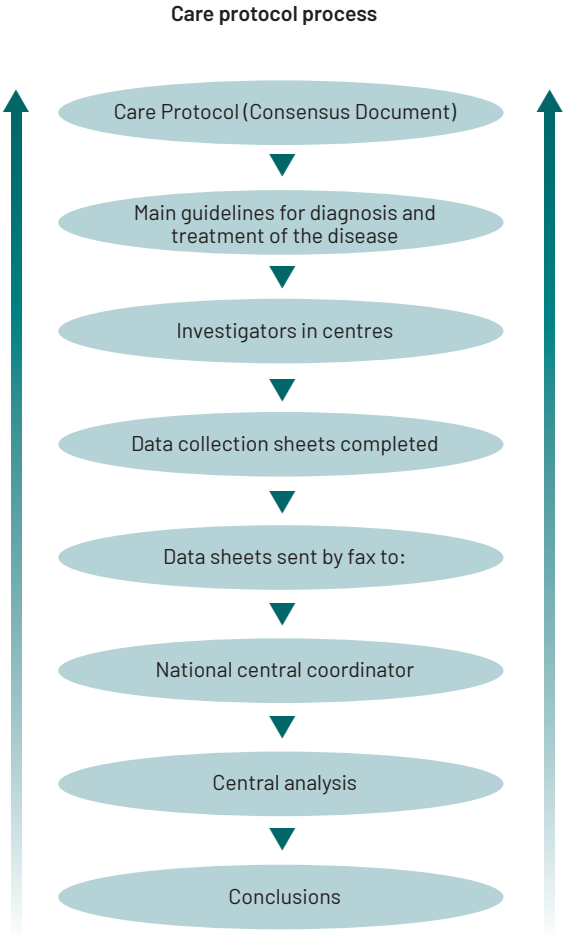


Figure 1. Key guidelines for disease diagnosis and treatment flowchart

Why is RWD crucial for personalised medicine in haematological malignancies?

RWD studies examine real-world conditions without discrimination or omission of patient type, collecting information from real clinical practice settings that are not pre-planned or pre-conditioned. They cannot replace the critical safety and efficacy data generated by clinical trials, but they can help to consolidate these data by allowing comparisons across a wide variety of real clinical settings.[2,5]

Thus, RWD research tends to provide more complete information on the safety and effectiveness of treatments in a wider variety of clinical settings.[5,6] They are therefore key to obtaining data on more hidden or subtle statistical trends that require large sample sizes (thousands or even hundreds of thousands), or much longer periods (several decades, even) to reveal.[7] This is especially necessary in haematological malignancies, where the constellation of biological variants makes it very difficult to obtain, in a clinical study setting, a sufficient sample of cases of a particular disease entity, such as a mutation or a specific set of them, for example. It is precisely in the study of haematological malignancies where a much greater abundance of information and cross-referencing of data is required due to the tremendous genetic and epigenetic phenomena that exist. On the other hand, RWD research is necessary in those types of research where it is not possible to set up more than a single arm, because they are serious and rare diseases where randomisation is often not feasible or ethically reprehensible, and where there is no choice but to rely on external control databases from clinical trials or previous RWD studies.[8]

Finally, RWD studies provide a type of information that can also be of great interest: patient-perceived outcomes.[9] This is particularly interesting in haematological malignancies, where the processes unfortunately often involve considerable morbidity.

Addressing the peculiarities of the main haematological malignancies – illnesses caused by the uncontrolled growth of bone marrow cells, which damage normal tissues and can lead to fatal bone marrow failure – we identified priority needs for RWD research, summarised as follows:

- **Monoclonal gammopathies (especially multiple myeloma):** These diseases are caused by the proliferation of a clone of mutated plasma cells (white blood cells responsible for antibody production) that secrete an abnormal protein (paraprotein). In multiple myeloma (MM), these cells grow without limits within the bones, leading to their destruction. In this field, RWD research is highly valuable for the complete and final evaluation of new therapies in patients with comorbidities who are often underrepresented in clinical trials. It also helps characterise predictive models of medium- and long-term response. This is of particular interest in smoldering myeloma (a relatively indolent hyperparaproteinemia that can sometimes progress to active myeloma) and in some specific genotypic profiles of MM.
- **Acute myeloblastic leukaemia (AML):** In this disease, the cellular proliferation originates from myeloid blood cells, and the critical damage, as in acute lymphoblastic leukaemia (ALL), is the eventual failure of the bone marrow. Here, RWD is crucial for a better understanding of prognostic values closely related to certain genetic mutations. Additionally, RWD can help define the optimal indication for hematopoietic stem cell transplantation for each case or clinical profile and for the exhaustive comparison of different treatments in very diverse patient populations.
- **Acute Lymphoblastic Leukaemia (ALL):** In this disease, the cellular proliferation originates from the lymphoid blood cells. RWD in this area is valuable for long follow-ups of patients with new targeted therapies and for comparing, across sufficient samples and different age strata, the various polychemotherapy regimens.
- **Bone marrow failure syndromes (BMFS):** The main disease in this group is aplastic anaemia, in which the bone marrow ceases to produce the daily quantity of cells necessary for the various functions of the blood. RWD is of interest here, above all, to obtain a sufficient database to study the response to immunosuppressive treatments and transplantation. This is especially important for disorders with such a low incidence, where gathering a sufficient sample to study any reality is always the greatest handicap.
- **Chronic myeloproliferative and lymphoproliferative processes:** These are less aggressive neoplastic proliferations of more mature cells from either myeloid or lymphoid lineages. RWD is of particular interest in assessing the impact of inhibitors in real-life settings, as well as in advancing the definition of more personalised treatments.
- **Lymphomas:** Cancer of the lymphatic nodes or spleen. RDW is highly valuable for long-term follow-up of patients treated with immunotherapy or emerging therapies in refractory conditions.

Across all these diseases, extensive biological research is crucial for discovering better, specific biomarkers to predict treatment response with greater precision.

Across all these diseases, extensive biological research is crucial for discovering better, specific biomarkers to predict treatment response with greater precision. This research also involves collecting more comprehensive information on idiosyncratic toxicities of new treatments or adverse reactions with delayed manifestations. In this latter aspect, RWD proves particularly useful for tracking second malignancies and for the complete study of certain more peculiar and unexpected toxicities. A good example of this is the novel CAR-T therapy, a modern, genetically-mediated immunological treatment that involves reprogramming the patient's lymphocytes by integrating new genetic information into their genome, enabling them to specifically recognise and attack malignant cells.

PETHEMA's experience so far at RWD

PETHEMA is active on several research fronts where RWD research is providing a decisive push towards a better understanding of diseases and their treatments. This work is carried out by the main study groups that make up PETHEMA, principally the Spanish Myeloma Group (GEM), the AML group and the ALL group. PETHEMA's experience to date in this area focuses mainly on AML, ALL, and MM, but includes interesting work in other areas like bone marrow failure.

PETHEMA's AML group holds a unique asset, allowing extensive RWD research and diverse analytical approaches. This refers to the AML epidemiological registry, which has stored all clinical data and much of the correlated biological data from over 24,600 patients diagnosed with AML, including about 5,300 with the acute promyelocytic leukaemia (APL) subtype, for several decades. Drawing on this vast

Close-up -blood cells – leukocytes erythrocytes bloodstream

Image: Freepik

number of cases (mostly thousands) of certain predefined clinical profiles, valuable published research has been conducted. These studies have described special patient populations,[11-16] and they've validated certain pharmacological treatments,[14,17,18,] even demonstrating advantages over others.[12,19,20-27] This research has also shown the suitability of allogeneic transplantation for AML treatment and analysed the incidence of secondary malignancies.[12,15,19,20-24,27] Furthermore, it's allowed for comparisons with historical external controls,[21,28,29] and importantly, it's demonstrated the prognostic value of Minimal Residual Disease (MRD) and its key role in risk stratification.[30,31] Above all, a huge amount of molecular data has been studied to better characterise clinical and genetic patterns linked to treatment responses and their associated prognoses.[13,14,15,18,22-26,28,29,31-34]

Research conducted within PETHEMA's Spanish Myeloma Group (or Grupo Español de Mieloma; GEM) is tremendously prolific through clinical trials. RWD research is less abundant but equally interesting, and has been carried out mainly through various observational and translational studies. At this point, we highlight studies carried out for the evaluation of certain special patient populations,[35] studies to verify the validity of certain pharmacological treatments or their superiority compared to others,[36,37] the characterisation of genetic profiles associated with different prognostic value[38,39] or other studies due to imminent or future publication that confirm the validity of monitoring MRD in MM for the optimisation of disease management (EMR-Clinical study, pending to be promptly published) or that search for factors predictive of transformation to MM, amyloidosis, Waldeström macroglobulinemia, or of complication to a severe infection (NoMoreMGUS).

In the area of ALL, the main RWD work is based on observational studies that have demonstrated the prognostic value of MRD in Phi-negative ALL patients, compared different treatment regimens, and measured the impact of allogeneic transplantation versus chemotherapy alone.[40,41] On the other hand, the ALL Group decided to collaborate with the Harmony Alliance in order to establish conclusions of prognostic value for a certain type of mutation,[43,44] or to analyse a huge amount of biological data to define genetic patterns associated with specific treatment responses and differing prognoses. The HARMONY Alliance is the European academic entity responsible for an initiative that collects extensive big data from several haematological malignancies. This data is obtained from numerous clinical trials and observational studies conducted by many European research groups. It is worth noting that several papers have already been published from this collaboration, with some involving the PETHEMA group.

In the field of bone marrow failure, an observational study published last year paradoxically revealed worse survival outcomes in patients with a moderate bone marrow aplasia profile compared to those with a severe or very severe profile.[45] There are also two soon-to-be-published ongoing observational studies (PIRE and APPRI-PNH), which will analyse the therapeutic outcomes of patients with paroxysmal nocturnal haemoglobinuria (PNH) treated with a different complement inhibitor in each study.

Problems of research with RWD

Extensive literature debates whether RWD research truly meets full quality standards, given that it may bypass certain rigours of Cartesian research developed through clinical trials.[46] What is certain, however, is that as with all research, its ultimate validity rests with the researchers. These individuals, aware of RWD's limitations and

adhering to necessary precautions, can achieve truly reliable results.[47]

Managing a huge amount of information is a key limitation of RWD research. The sheer scale and diversity of this data require powerful and complex computer equipment, which presents a substantial problem due to the high costs of both using and maintaining the necessary infrastructure.[48] These costs are mainly related to the need for continuous review and updating of data and data quality, modifications to the source system, changes in system specifications, and identifying variations in implementation between sites to address the challenges of incompatibilities that arise between different information and analysis systems. Another limitation is the fact that some differences between data from different sites cannot be resolved by a standardised data model and require a close level of cooperation with site staff to overcome difficulties in unambiguously interpreting information. Confounding factors are a significant challenge when comparing the effectiveness and safety of treatments using real-world data, particularly due to the data's diverse nature. Rigorous control of design and analytical fit is required to obtain consistent and truthful estimates that withstand any test of irrefutability.[49]

Regarding ethical issues, the growing problem of strict, regulated, explicit informed consent in RWD research warrants attention, particularly where the routine use of data must be intensive and fluid.[50] Similarly, rigorous data pseudo-anonymisation policies, while attempting to protect privacy at all costs, sometimes exceed what is reasonable. This hinders the rapid and efficient obtaining of sufficient quality information in complicated clinical or highly relevant public health issues.[51]

Reflections on the future

The refinement of clinical research and the trend towards maximum sample universalisation (aiming not to lose a single study-worthy case) will exponentially increase the complexity of work systems. These systems will, in turn, become increasingly dependent on technological advances. The use of artificial intelligence is already a fact and will continue to gain ground here, as in many other areas of scientific activity. This aligns with the need to acquire the enormous capacity to analyse highly specific and often hidden biological data, detecting their logical clinical and biological interrelationships. This will allow conclusions to be drawn that are impossible with conventional statistical analyses.

The widespread use of nano-robotics, which is already being explored for certain therapeutic purposes, may also have an exciting role to play in the precise and immediate collection of intracorporeal intimate cellular and biological niche data.

Progress in improving clinical (patient-centred) and biological (lab work) diagnostic tools and techniques is essential. We need to make them more reliable, faster, more efficient, and less invasive, not only to advance research but also to improve medical practice. Additionally, further work is needed to improve patient compliance with wearable devices by making them more discreet, efficient, and less disruptive to daily life.

Finally, in the defence of scientific freedom, we must offer a constructive critique of regulatory issues. This critique is presented as the authors' personal opinion, with the sole aim of stimulating healthy reflection on how to improve scientific progress. Certain issues in this respect should provoke a paradigm shift in the legislator's approach to facilitate the transition towards a more scientifically productive future that is ultimately pragmatically useful to humanity.

Data protection legislation seeks to protect fundamental rights, but it can significantly impede the dynamism in the exercise of scientific freedom. This freedom, it is crucial to remember, always seeks the greater common good of scientific progress and the advancement of human well-being. A new balance in this game of defence between individual and collective rights needs to be fostered in this area. This would allow the global medical-scientific research system to function better without compromising fundamental protections.[52] Science consistently outpaces the law, so there is an imperative need for regulators and legislators involved in the control of health research activity to urgently adapt to the needs of scientific discovery. They need to shape new principles to harmonise the most appropriate approach to operational efficiency, ethics, and governance globally.[52,53]

Acknowledgements

These acknowledgements are offered by the first author. I wish to express my heartfelt thanks to Dr Joaquín Díaz Mediavilla, my mentor and friend, for his guidance, integrity, and unwavering support. I am also deeply grateful to Dr Evarist Feliu Frasnedo for his lifelong encouragement and inspiration, and to Dr Juan José Lahuerta for his loyalty and example of professional commitment. I thank my PETHEMA colleagues for over two decades of collaborative scientific progress, and the coauthors of this paper for their contributions. Finally, I extend sincere appreciation to the Foundation team – exceptional professionals who make our work possible.

Disclaimers

The opinions expressed in this article are the authors' own and not necessarily shared by their employer, the PETHEMA members, or EMWA.

Disclosures and conflicts of interest

The authors declare no conflicts of interest.

References

1. EU Trials Tracker [accessed 2025 Jun 2]. Available from: https://eu.trialstracker.net/?all
2. Garrison LP, Neumann PJ, Erickson P, et al. Using real-world data for coverage and payment decisions: the ISPOR Real-World Data Task Force report. Value Health. 2007;10(5):326-35. doi:10.1111/j.1524-4733.2007.00186.x
3. FDA. Real-World Evidence. Silver Spring (MD): FDA; 2024 May 16 [cited 2025 Jun 2]. Available from: https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence
4. Ford I, Norrie J. Pragmatic trials. N Engl J Med. 2016;375(5):454-63. doi:10.1056/NEJMra1510059
5. Berger ML, Sox H, Willke RJ, et al. Good practices for real-world data studies of treatment and/or comparative effectiveness: Recommendations from the joint ISPOR–ISPE Special Task Force on real-world evidence in health care decision making. Pharmacoepidemiol Drug Saf. 2017;26(9):1033-9. doi:10.1002/pds.4297
6. Greene JA, Podolsky SH. Reform, regulation, and pharmaceuticals – The Kefauver–Harris Amendments at 50. N Engl J Med. 2012;367(16):1481-3. doi:10.1056/NEJMp1210007

7. The Association of the British Pharmaceutical Industry. The vision for real-world data: Harnessing the opportunities in the UK. Demonstrating value with real-world data. London: ABPI; 2011.
8. Ross TR, Ng D, Brown JS, et al. The HMO Research Network Virtual Data Warehouse: A public data model to support collaboration. EGEMS (Wash DC). 2014;2(1):1049. doi:10.13063/2327-9214.1049
9. Glasgow RE, Magid DJ, Beck A. Practical clinical trials for translating research to practice: design and measurement recommendations. Med Care. 2005;43(6):551-7. doi:10.1097/01.mlr.0000163645.41407.09
10. López-Briz E, Ordovás-Baines JP. Del real-world data a la farmacoterapia personalizada. ¿Por dónde vamos? Farm Hosp. 2021;45(Suppl 1):S1-2.
11. Sossa-Melo C, Abello-Polo V, Salazar LA, et al. Characteristics, outcomes and treatment patterns in acute myeloid leukemia patients 60 years or older in Colombia: a RENEHOC-PETHEMA study. Ann Hematol. 2025;104(1):369-81. doi:10.1007/s00277-024-06120-0
12. Hernández-Boluda JC, Martínez-Cuadrón D, Pereira A, et al. Acute leukemia arising from myeloproliferative or myelodysplastic/myeloproliferative neoplasms: A series of 372 patients from the PETHEMA AML registry. Leuk Res. 2022;115:106821. doi:10.1016/j.leukres.2022.106821
13. Palanques-Pastor T, Megías-Vericat JE, Martínez P, et al. Characteristics, clinical outcomes, and risk factors of SARS-COV-2 infection in adult acute myeloid leukemia patients: experience of the PETHEMA group. Leuk Lymphoma. 2021;62(12):2928-38. doi:10.1080/10428194.2021.1948031
14. Martínez-Cuadrón D, Serrano J, Gil C, et al. Evolving treatment patterns and outcomes in older patients (≥60 years) with AML: changing everything to change nothing? Leukemia. 2021;35(6):1571-85. doi:10.1038/s41375-020-01058-4
15. Sobas M, Talarn-Forcadell MC, Martínez-Cuadrón D, et al. PLZF-RARα, NPM1-RARα, and other acute promyelocytic leukemia variants: the PETHEMA registry experience and systematic literature review. Cancers (Basel). 2020;12(5):1313. doi:10.3390/cancers12051313
16. Salamero O, Martínez-Cuadrón D, Sobas M, et al. Real-life outcomes of patients aged ≥75 years old with acute promyelocytic leukemia: experience of the PETHEMA registry. Leuk Lymphoma. 2019;60(11):2720-32. doi:10.1080/10428194.2019.1607327
17. Vives S, Quintela D, Morgades M, et al. Salvage therapy with second-generation inhibitors for FLT3-mutated acute myeloid leukemia: a real-world study by the CETLAM and PETHE MA groups. Cancers (Basel). 2024;16(23):4028. doi:10.3390/cancers16234028
18. Sobas M, Knopinska-Posluszny W, Piątkowska-Jakubas B, et al. Incidence, risk factors, and outcomes of second neoplasms in patients with acute promyelocytic leukemia: the PETHEMA-PALG experience. Ann Hematol. 2024;103(2):451-61. doi:10.1007/s00277-023-05582-y
19. Serrano J, Martínez-Cuadrón D, Gil C, et al. Autologous or allogeneic hematopoietic stem cell transplantation as front-line treatment for adult secondary acute myeloid leukemia patients: the PETHEMA registry experience. Transplant Cell Ther. 2025;S2666-6367(25)01046-2. doi:10.1016/j.jtct.2025.02.011
20. Lloret-Madrid P, Boluda B, Martínez-López J, et al. Treatment patterns and outcomes in secondary acute myeloid leukemia arising after hypomethylating agents: PETHEMA registry study. Cancer. 2025;131(1):e35696. dpo" doi:10.1002/cncr.35696
21. Bernal T, Moreno AF, de LaIglesia A, et al. Clinical outcomes after CPX-351 in patients with high-risk acute myeloid leukemia: A comparison with a matched cohort from the Spanish PETHEMA registry. Cancer Med. 2023;12(14):14892-901. doi:10.1002/cam4.6120
22. Martínez-Cuadrón D, Megías-Vericat JE, Gil C, et al. Outcomes after intensive chemotherapy for secondary and myeloid-related changes acute myeloid leukemia patients aged 60 to 75 years old: a retrospective analysis from the PETHEMA registry. Haematologica. 2024;109(1):115-28. doi:10.3324/haematol.2022.282506
23. Ayala R, Carreño-Tarragona G, Barragán E, et al. Impact of FLT3-ITD mutation status and its ratio in a cohort of 2901 patients undergoing upfront intensive chemotherapy: a PETHEMA registry study. Cancers (Basel). 2022;14(23):5799. doi:10.3390/cancers14235799
24. Martínez-Cuadrón D, Serrano J, Mariz J, et al. Characteristics and outcomes of adult patients in the PETHEMA registry with relapsed or refractory FLT3-ITD mutation-positive acute myeloid leukemia. Cancers (Basel). 2022;14(11):2817. doi:10.3390/cancers14112817
25. Labrador J, Martínez-Cuadrón D, de la Fuente A, et al: On behalf of Pethema Group. Azacitidine vs. decitabine in unfit newly diagnosed acute myeloid leukemia patients: results from the PETHEMA registry. Cancers (Basel). 2022;14(9):2342.. doi:10.3390/cancers14092342
26. Labrador J, Saiz-Rodríguez M, de Miguel D, et al. Use of venetoclax in patients with relapsed or refractory acute myeloid leukemia: the PETHEMA registry experience. Cancers (Basel). 2022;14(7):1734. doi:10.3390/cancers14071734
27. Rodríguez-Arbolí E, Martínez-Cuadrón D, Rodríguez-Veiga R, et al. Long-term outcomes after autologous versus allogeneic stem cell transplantation in molecularly-stratified patients with intermediate cytogenetic risk acute myeloid leukemia: a PETHEMA study. Transplant Cell Ther. 2021;27(4):311.e1-10. doi:10.1016/j.jtct.2020.12.029
28. Bérard E, Röllig C, Bertoli S, et al. A scoring system for AML patients aged 70 years or older, eligible for intensive chemotherapy: a study based on a large European data set using the DATAML, SAL, and PETHEMA registries. Blood Cancer J. 2022;12(7):107. doi:10.1038/s41408-022-00700-x
29. Martínez-Cuadrón D, Megías-Vericat JE, Serrano J, et al. Treatment patterns and outcomes of 2310 patients with secondary acute myeloid leukemia: a PETHEMA registry study. Blood Adv. 2022;6(4):1278-95. doi:10.1182/bloodadvances.2021005335

30. Paiva B, Vidriales MB, Sempere A, et al. Impact of measurable residual disease by decentralized flow cytometry: a PETHEMA real-world study in 1076 patients with acute myeloid leukemia. Leukemia. 2021;35(8):2358-70. doi:10.1038/s41375-021-01126-3
31. Sargas C, Ayala R, Larráyoz MJ, et al. On behalf of the Pethema Group. Molecular landscape and validation of new genomic classification in 2668 adult AML patients: real-life data from the PETHEMA registry. Cancers (Basel). 2023;15(2):438. doi:10.3390/cancers15020438
32. Labrador J, Martínez-Cuadrón D, Boluda B, et al. Evolving patterns and clinical outcome of genetic studies performed at diagnosis in acute myeloid leukemia patients: real-life data from the PETH EMA registry. Cancer. 2024;130(20):3436-51. doi:10.1002/cncr.35431
33. Sargas C, Ayala R, Larráyoz MJ, et al, Pethama Cooperative Study Group Comparison of the 2022 and 2017 European LeukemiaNet risk classifications in a real-life cohort of the PETHEMA group. Blood Cancer J. 2023;13(1):77. doi:10.1038/s41408-023-00835-5
34. Sitges M, Boluda B, Garrido A, et al. Acute myeloid leukemia with inv(3) (q21.3q26.2)/t(3;3)(q21.3;q26.2): study of 61 patients treated with intensive protocols. Eur J Haematol. 2020;105(2):138-47. doi:10.1111/ejh.13417
35. Martínez-López J, Mateos MV, Encinas C, et al. Multiple myeloma and SARS-CoV-2 infection: clinical characteristics and prognostic factors of inpatient mortality. Blood Cancer J. 2020;10(10):103. doi:10.1038/s41408-020-00372-5
36. de la Rubia J, Alonso R, Clavero ME, et al. Belantamab mafodotin in patients with relapsed/refractory multiple myeloma: results of the compassionate use or the expanded access program in Spain. Cancers (Basel). 2023;15(11):2964. doi:10.3390/cancers15112964
37. Rodriguez-Otero P, Sirvent M, González-Rodríguez AP, et al. Pomalidomide, cyclophosphamide, and dexamethasone for the treatment of relapsed/refractory multiple myeloma: real-world analysis of the Pethema-GEM experience. Clin Lymphoma Myeloma Leuk. 2021;21(6):413-420. doi:10.1016/j.clml.2021.02.004
38. Garcés JJ, Bretones G, Burgos L, et al. Circulating tumor cells for comprehensive and multiregional non-invasive genetic characterization of multiple myeloma. Leukemia. 2020;34(11):3007-18. doi:10.1038/s41375-020-0883-0
39. Termini R, Žihala D, Terpos E, et al. Circulating tumor and immune cells for minimally invasive risk stratification of smoldering multiple myeloma. Clin Cancer Res. 2022 Nov;28(21):4771–81. doi:10.1158/1078-0432.CCR-22-1594
40. Ribera JM, Oriol A, Morgades M, et al. Treatment of high-risk Philadelphia chromosome-negative acute lymphoblastic leukemia in adolescents and adults according to early cytologic response and minimal residual disease after consolidation assessed by flow cytometry: final results of the PETHEMA ALL-AR-03 trial. J Clin Oncol.2014;32(15):1595–604. doi:10.1200/JCO.2013.52.2425

Authors information

Alfonso J. Santiago-Marí, MD. Fundación PETHEMA, Madrid, Spain. Haematologist, political scientist, and executive director of the PETHEMA Foundation, Dr Santiago-Mari has been an investigator in leukaemia clinical trials, a scientific communicator, and a medical writer. He has published several works on research activity in medicine and topics related to medicine and law.

Irene Salcedo-Marín, MA. Fundación PETHEMA, Madrid, Spain. An English philologist and Fine Arts graduate with over 20 years of experience in clinical trials, Salcedo-Marin has held various roles in pharmaceutical companies and CROs. She is currently Associate Manager at the PETHEMA Foundation.

Roberto Maldonado, MsC. Fundación PETHEMA, Madrid, Spain. Biologist, biochemist, and Clinical Operations Manager at the PETHEMA Foundation, Maldonado holds several research-focused Master's degrees and leads research quality oversight for GEM (Spanish Myeloma Group) projects within PETHEMA.

41. Ribera JM, García O, Gil C, et al. Pethema Group Comparison of intensive, pediatric-inspired therapy with non-intensive therapy in older adults aged 55–65 years with Philadelphia chromosome-negative acute lymphoblastic leukemia. Leuk Res. 2018;68:79–84. doi:10.1016/j.leukres.2018.03.010
42. Ribera JM, Morgades M, Ciudad J, et al. Chemotherapy or allogeneic transplantation in high-risk Philadelphia chromosome-negative adult lymphoblastic leukemia. Blood. 2021 Apr 8;137(14):1879–94. doi:10.1182/blood.2020007311
43. HARMONY Alliance. ALL-2: Prognostic value of IKZF1 gene alternations in adult BCP-ALL [Internet]. HARMONY Alliance; [cited 2025 Jun 2]. Available from: https://www.harmony-alliance.eu/ projects/ research-project/all-2-prognostic-value-of-ikzf1-gene-alternations-in-adult-bcp-all
44. HARMONY Alliance. Use of Big Data to improve outcomes for patients with acute lymphoblastic leukemia (ALL) – 2019. HARMONY Alliance [cited 2025 Jun 2]. Available from: https://www.harmony-alliance.eu/projects/research-project/use-of-big-data-to-improve-outcomes-for-patients-with-acute-lymphoblastic-leukemia-all-2019
45. Vallejo C, Rosell A, Xicoy B, et al. A multicentre ambispective observational study into the incidence and clinical management of aplastic anaemia in Spain (IMAS study). Ann Hematol. 2024;103(3):705-713. doi:10.1007/s00277-023-05602-x
46. Duffy DJ. Problems, challenges and promises: perspectives on precision medicine. Brief Bioinform. 2016;17(3): 494-504. doi:10.1093/bib/bbv060
47. Willke RJ, Mullins CD. "Ten commandments" for conducting comparative effectiveness research using real-world data. J Manag Care Pharm. 2011;17(8 Suppl):S10-5. doi:10.18553/jmcp.2011.17.s9-a.S10
48. Sivananthan A. Systems biology and real-world data as drivers of change in drug research and development. Med Writ. 2023;32(1).. doi:10.56012/awvf3242
49. Ross TR, Ng D, Brown JS, et al. The HMO Research Network Virtual Data Warehouse: a public data model to support collaboration. EGEMS (Wash DC). 2014;2(1):1049. doi:10.13063/2327-9214.1049
50. Curtis LH, Weiner MG, Boudreau DM, et al. Design considerations, architecture, and use of the Mini-Sentinel distributed data system. Pharmacoepidemiol Drug Saf. 2012;21 Suppl 1:23-31. doi:10.1002/pds.2336
51. Council for International Organizations of Medical Sciences (CIOMS). Real-world data and real-world evidence in regulatory decision making. Geneva: CIOMS; 2024. doi:10.56759/kfxh6213
52. Rawlins MD. Spontaneous reporting of adverse drug reactions. I: The data. Br J Clin Pharmacol. 1988;26(1):1-5. doi:10.1111/j.1365-2125.1988.tb03356.x
53. García López JL, Llanos Señarís JE, et al. Aportación de los "Real-World Data (RWD)" a la mejora de la práctica clínica y del consumo de recursos de los pacientes. Madrid: Fundación Gaspar Casal; 2014.

Carmen López-Carrero, MPharm Fundación PETHEMA, Madrid, Spain. Pharmacist and Projects Director at the PETHEMA Foundation, Lopez-Carrero is coordinating over 60 studies. She oversees all stakeholder activities across the projects and brings extensive clinical research experience from previous roles in pharmaceutical companies.

David Martínez-Cuadrón, MD, Haematology Department, Hospital Universitari i Politècnic La Fe. València, Spain. A haematologist and Assistant Coordinator of the international PETHEMA group at University Hospital La Fe in Valencia, Spain, Martinez-Cuadron has contributed to numerous high-impact scientific publications, participated in over 150 clinical trials – most in early development phases – and authored more than 100 presentations at national and international congresses.

Pau Montesinos, PhD, Haematology Department, Hospital Universitari i Politècnic La Fe. València, Spain. Attending haematologist and head of the Leukaemia Unit at the Department of Haematology, University Hospital La Fe in Valencia, Montesinos is principal investigator in over 100 trials (phases 1 to 3) in acute leukaemias. Chair of the international PETHEMA AML/APL group, he is also coordinator of an epidemiological registry of Iberian and Latin American AML/APL patients.

Real-world data in clinical development: Statistical considerations and reporting challenges

Natasa Rajicic[1], **Anders Mattsson**[2], **Maria Kołtowska-Häggström**[3]
[1] SC Rules, LLC, Warwick, NY, USA
[2] Stockholm, Sweden
[3] Proper Medical Writing, Warsaw, Poland

doi: 10.56012/kygt4232

Correspondence to:
Natasa Rajicic
natasa.rajicic@outlook.com

Abstract
Real-world data have an increasingly important role in clinical development and regulatory decision-making. When incorporated correctly, they can provide a unique and valuable insight into patient populations, treatment patterns, and health outcomes in support to the traditional clinical development. To that end, transparency in reporting, including clear documentation of study populations, data sources, statistical methods, and limitations, is critical, particularly when seeking regulatory acceptance. Recognised standards of reporting should be considered as they can help to enhances reproducibility and regulatory acceptance.

Introduction
Real-world data (RWD) and real-world evidence (RWE) have long been utilised for a variety of purposes such as characterisation of population health and disease trends or to study risk associated with different exposures, just to name a few. RWD and RWE have also been an important part of drug safety surveillance, especially following a market drug approval after which a new medicine starts to be used in clinical practice. More recently, RWD and RWE are increasingly used in clinical development and regulatory decision-making in new drug applications. While the conventional clinical trials, randomised clinical trials (RCTs) in particular, have long been a cornerstone of clinical development programmes and regulatory submissions due to their rigorous designs that enable causal interpretations, real-world data is becoming an important supplementary source of evidence in clinical development of new medicines and regulatory approval decisions.

The term *RWD* refers to data derived from sources that are outside of the conventional clinical trials, including, for example, electronic health records (EHRs), medical claims databases, patient registries, and wearable health technologies.[1] Evidence generated from RWD studies has been used to inform disease histories, safety surveillance in post-marketing, quality of life outcomes, or treatment effectiveness in clinical practice, just to name a few.[2,3] Recently, the integration of RWD into clinical development has been recognised and promoted by regulatory agencies such as the US Food and Drug Administration (FDA) and the European Medicines Agency (EMA), for their potential to support and inform drug approvals and policy decisions.[4,5]

While any research design is concerned with issues such as selection bias, confounding, and variabilities in data collection and endpoint definitions, these concerns are even more prominent when it comes to RWD. Anticipating and addressing these issues is crucial for maintaining the integrity, validity, and applicability of findings that result from RWD. Transparent reporting contributes to the credibility and validity of findings, particularly when RWD are used as a part of regulatory submissions.

Transparent reporting contributes to the credibility and validity of findings, particularly when RWD are used as a part of regulatory submissions.

This article provides an overview of some common statistical methodologies employed when analysing real-world data, and discusses the challenges associated with reporting RWD findings, with a particular emphasis on statistical and interpretation issues essential for maintaining methodological integrity, validity, and transparency when reporting analyses of RWD.

Why RWD in clinical development
RWD have long been used to support post-marketing safety surveillance, continued benefit-risk evaluations, and label extension applications, e.g., in rare diseases that have limited patient populations or in situations where traditional clinical trials would be unfeasible or unethical. An increasingly attractive use of RWD is within the clinical development phase, where for example, a traditional control group would be impractical or unethical. Here, RWD are used as a source for creating an external control group, thus enabling a structured comparator where otherwise one would be absent from the investigation.

In fact, several features of traditional clinical trials, and RCTs in particular, make the use of RWD an attractive fit complementing clinical development. For example, clinical trials often have strict inclusion and exclusion criteria as these can help create a more homogeneous study population that can in turn reduce overall variability and increase precision and power of estimation. Homogeneity of the study population can also reduce the impact of known and unknown confounding variables. However, strict patient selection criteria can make the study interpretation less generalisable to real-world clinical settings. Rare disease studies with limited patient population pools often face challenges in enrolling enough participants for an adequately designed and powered study. Rigorously designed, monitored, and executed studies are often prohibitively expensive and essentially impossible to carry out unless conducted by large pharmaceutical sponsors or consortia. Interventions that are studied in highly controlled clinical trials that do not mirror routine clinical practice limit the generalisability of the results.

Image: Freepik

Finally, clinical trials may not be conducted over a long enough period to provide data on either late-emerging adverse events or effects in incurable chronic conditions that require life-long treatment.

Given these limitations, integrating RWD alongside RCTs can enhance evidence generation by providing insights into broader patient populations, long-term effects, and real-world treatment effectiveness.[6] Therefore, appropriately designed and executed analyses based on RWD can be a complementary and useful tool in filling the gaps present in traditional clinical trials.

Commonly used statistical methodologies in RWD analysis

Analyses and inferences based on RWD often require different statistical methods compared to the analytical approaches used in analyses of typical clinical trial, especially when contrasted with RCTs. This is because most statistical methods used in standard analyses assume that the patient groups being compared are reasonably well balanced – both in terms of known and unknown potential confounders – prior to the introduction of the intervention. They also assume that follow-up of participants remains comparable across groups, except for differences directly attributable to the intervention itself. Use of randomisation can help with the first issue, and principled adherence to a well-developed protocol that strives for controlled and uniform follow-up procedures can help deal with the second issue. Consequently, results of the statistical tests evaluating the differences among study groups can be potentially interpreted as causality.

The absence of the above two and other considerations when conducting the analyses based on RWD, as a result, necessitate application of statistical method that are sensitive to such issues. The following is a summary of few statistical approaches that were primarily developed to handle data outside of RCTs and that can be found useful when analysing RWD.

Regression analysis models

Regression analysis is a key statistical tool used to explore and quantify relationships between variables, such as treatment exposure and clinical outcomes. In epidemiological studies and RWD analyses, regression models such as linear and logistic regression, repeated measures analysis, or Cox proportional hazards modes, are commonly applied to control for confounding factors when characterising relationships between exposure and responses. By adjusting for patient demographics, comorbidities, and other covariates, regression can help isolate the impact of a specific variable of interest, enhancing the validity of conclusions drawn from non-randomised, real-world settings. However, regression analysis in RWD has limitations, including confounding due to unmeasured or misclassified variables, model misspecification, selection bias, or missing data. Nevertheless, regression methods, especially when used properly, remain an essential and widely used tool for the statistical analyses in RWD settings.

By adjusting for patient demographics, comorbidities, and other covariates, regression can help isolate the impact of a specific variable of interest.

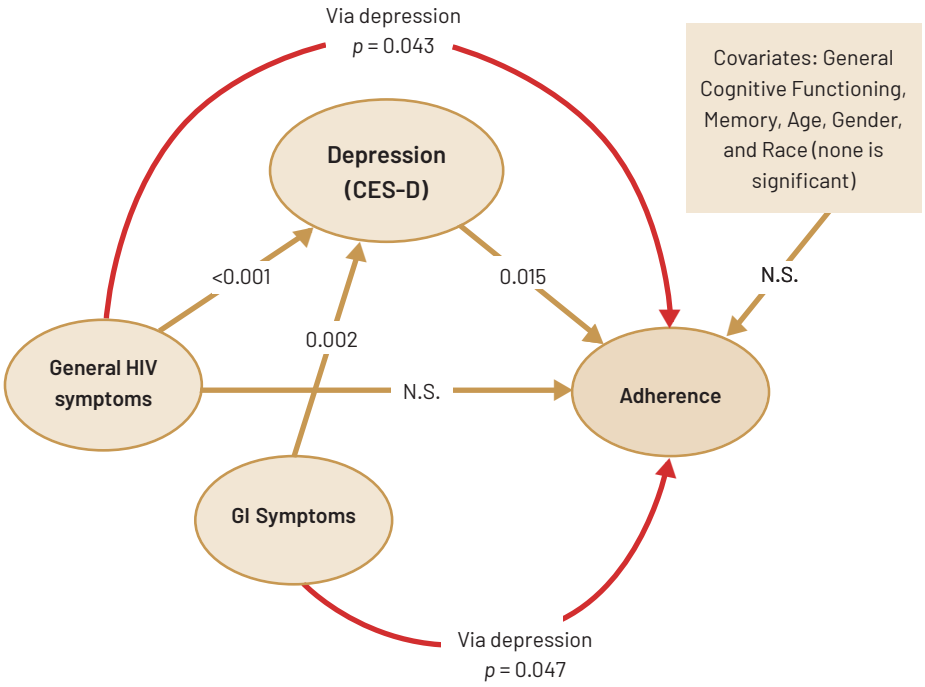


Figure 1. Graphical illustration of use of structural equation models to characterise and evaluate causal relationships involving direct and indirect effects among variables

The graph indicates that both general HIV-related symptoms and gastrointestinal (GI) symptoms are directly associated with higher levels of depressive symptoms. In turn, higher levels of depressive symptoms are directly linked to lower medication adherence. Notably, general HIV-related symptoms do not have a direct effect on adherence; rather, their impact is indirect, mediated by depressive symptoms. This suggests that an increase in general HIV-related symptoms is associated with increased depressive symptoms, which subsequently lead to poorer adherence. Similarly, GI symptoms also exert an indirect effect on adherence through depressive symptoms, with no direct relationship observed between GI symptoms and adherence.

Propensity score methods

Propensity score methods, such as propensity score matching and inverse probability weighting, were developed to help create groups of patients that are balanced with respect to observed baseline characteristics in observational studies.[7] Propensity score methods can generally be divided into two categories: propensity score matching, which attempts to pair patients with similar characteristics across different treatment arms, and inverse probability of treatment weighting, which assigns weights to patients based on their propensity scores to create a pseudo-randomised population. In this way, propensity scores construct pools of patients that appear similar with respect to the distributions of covariates, irrespective of the actual treatment subsequently received. Thus, if properly executed, analysis of the differences in outcomes between treatment groups that incorporate propensity scores could help adequately evaluate treatment differences in RWD analyses.

Causal inference methods

The topic of causal inference is rich and long-standing. Numerous methods and relevant theories have been developed.[8] Moreover, causal inference methods are rarely a standard topic in statistical academic programmes, even in advanced post-graduate studies, and like many advanced methodologies, they should be handled by experienced professionals only. Many of these approaches aim to examine relationships between variables or concepts that are not directly observable from the data or attempt to estimate causal relationships from data in the presence of various types of confounding. Here we introduce only a few methods as an illustration; more comprehensive reviews can be found elsewhere.[8,9]

Structural equation modelling (SEM) provides a framework for examining relationships between observed variables and underlying constructs – latent variables – that cannot be directly observed (e.g., depression or quality of life) but are inferred from other measurable variables. SEMs utilise and combine methods of factor analysis and regression and can be visualised using diagrams that depict hypothesised causal directional paths among variables (Figure 1). An example of SEM application is a study to examine the process by which direct and indirect effects of HIV-related symptoms are related to adherence to antiretroviral therapy as well as whether the symptom of depression acts as a mediator of this relationship.[10]

Bayesian methods

Bayesian statistical methods, in which inference is made based on data-driven updates to prior beliefs, has found numerous applications in the design and analysis of clinical trial data. Bayesian analysis approaches incorporate prior external information, for example evidence from completed clinical trials or expert opinion, into current analyses, enabling more robust inference even when data are limited or heterogeneous. In RWD applications, Bayesian models can account for missing data by predicting unknown values using the available data (e.g. through multiple imputation), adjust for confounding factors through Bayesian propensity score methods, or account for variability across different populations using hierarchical modelling, as in *pragmatic* trials.[11] Bayesian analysis can also facilitate dynamic updating of inferences as new data become available, making them particularly valuable for ongoing studies and real-time decision-making.[12]

Table 1. Reporting strategies for ensuring transparency

Reporting topic	What needs to be described or included?	Why?
Information on data sources and its quality	• Data origin that is sufficiently and clearly detailed, with specific sources named (e.g., EHRs, claims, registries, etc.) and whether data were collected for a specific purpose or extracted from a database that collected data without a prespecified purpose • Details on how data were extracted or collected, managed, cleaned, and processed, including what steps were taken to maintain data integrity and quality	• Details around the processes of data collection and curation can provide important insights into any limitations or potential biases inherent to the data • A reader should be sufficiently informed prior to making decisions and/or interpretations based on the results
Sources of patient population	• Inclusion and exclusion criteria with attention to any specific characteristics • For registries, selection criteria described separately for the entire registry and for the subset of patients analysed in a specific registry-based study • When RWD are used to supplement data from clinical trials, detailed differences in the populations • Clear distinction between e.g., mining of the entire registry vs. targeted selection of data from a registry based on pre-specified inclusion/exclusion criteria	• Understanding the source population helps reviewers understand the generalisability of the findings and how representative results are of any targeted populations • Without randomised assignment to treatment, participants who are treated may be inherently and systematically different from those who are not
Analysis plans and methods	• Comprehensive description of the statistical methods used, all assumptions clearly stated • Justification of methodological choices and any alternative strategies considered • Report of sensitivity analyses conducted to assess robustness of findings	• Transparency regarding analytical methods used ensures reproducibility, a basic tenant of rigorous research • Regulatory agencies require transparency in how RWE studies are conducted
Sources of bias and confounding	• Potential sources of bias, such as selection bias and confounding, and how adjustments were made to minimise these biases • Results presented both before and after adjustment to illustrate their impact	• The source of the RWD is in routine clinical practice where factors like disease history, prior treatments, and clinical settings can influence outcomes and introduce bias when interpreting treatment effectiveness
Consistency in endpoint definitions	• Clinical endpoint definitions standardised across different data sources, including sites and institutions • Criteria used for endpoint derivation and the validation process	• Different RWD sources may define clinical endpoints differently which can translate into a different outcome or endpoint when it comes to the analysis • The degree of consistency in endpoint definitions is essential for interpretation of the results

Machine learning methods

Finally, in this brief overview of statistical methods, machine learning should be mentioned as well since RWD often involve large, complex datasets where machine learning can be particularly useful. The results include improved insight through supervised learning (e.g., Random Forests) or assistance in predictive modelling and treatment effect estimation.[13] Similarly, unsupervised learning such as clustering, or dimensionality reduction (e.g., principal component analysis or lasso regression) can help identify patterns within patient populations.[14]

Reporting of RWD

Ensuring transparency in reporting

Transparency is vital for ensuring credibility, reproducibility, and ultimately regulatory acceptance of analyses based on RWD. Table 1 highlights several important objectives when summarising evidence arising from RWD.

In addition to the statistical reporting issues listed in Table 1, a transparent reporting of RWD should include topics of missing data. Namely, these include quantification of the extent of missing data and its potential impact on interpretation of analysis results, methods used to handle missing data in the analysis (e.g., only complete cases analysed, or type of imputation method employed), as well as any sensitivity analyses completed to explore the influence of missing data assumptions on analysis results. In addition, whenever possible, access to study protocols and analytical code should be provided as this greatly enhance reproducibility. Open-source platforms are a good place for sharing, provided they adequately safeguard patient privacy.

> Bayesian analysis approaches incorporate prior external information, for example evidence from completed clinical trials or expert opinion, into current analyses, enabling more robust inference even when data are limited or heterogeneous

Importance of objective communication of findings

Like any study or data reporting, objective reporting and interpretation of findings is essential. (See, for example, the ClinicalTrials.gov repository.)[15] Given the inherent difficulties in establishing causal inference in results based on RWD, the importance of careful reporting of RWD analyses should be emphasised. A careful consideration not to overstate causal relationships, especially given the observational nature of RWD studies, is essential. Communication of RWD findings requires measured and balanced language with ample discussion of potential biases, while acknowledging any limitations. Remaining uncertainties should be highlighted, and, when meaningful, alternative explanations to the findings presented.[16]

Enhancing transparency not only strengthens confidence in RWE but also facilitates its successful integration into clinical decision-making and regulatory assessments. This not only builds confidence among regulators and the scientific community but also paves the way for the broader integration of RWD in healthcare.

Conclusions

RWD will increasingly be used to complement traditional clinical trial data. Their applications – from constructing external control arms and enhancing safety surveillance to supporting research in rare diseases, underscore their growing role in clinical development and regulatory decision-making. However, the inherent challenges of RWD, such as selection bias, confounding, and inconsistencies in data collection, require specialised analytical approaches such as propensity score methods, Bayesian techniques, and causal inference models. Additionally, transparency in reporting, including clear documentation of study populations, data sources, statistical methods, and limitations, is critical to maintaining scientific rigor, particularly when seeking regulatory acceptance. Recognised standards of reporting, such as those outlined by STROBE and RECORD,[17,18] should be considered as they can help to enhance reproducibility and regulatory acceptance.

This article has focused on statistical and reporting considerations, but numerous other aspects of RWE should be considered, including appropriate regulatory frameworks. Fortunately, significant strides in formalising common practices and providing industry guidance have already been accomplished.[4,5] Other important considerations that need addressing are data privacy, and ethical and security concerns when utilising RWD.[6] The growing collaboration across industry, academia, and regulatory bodies is encouraging and welcomed and will likely lead to industry-wide, recognised best practices towards greater utilisation of the RWD and RWE.

Image: Freepik

Disclosures and conflicts of interest

The authors declare no conflicts of interest.

References

1. Liu F, Panagiotakos D. Real-world data: brief review of the methods, applications, challenges and opportunities. BMC Med Res Methodol. 2022;22(1):287. doi:10.1186/s12874-022-01768-6
2. Dang A. Real-world evidence: A primer. Pharmaceut Med. 2023;37(1):25–36. doi:10.1007/s40290-022-00456-6
3. Corrigan-Curay J, Sacks L, Woodcock J. Real-world evidence and real-world data for evaluating drug safety and effectiveness. JAMA. 2018;320(9):867–8. doi:10.1001/jama.2018.10136
4. US Food and Drug Administration. Use of real-world evidence to support regulatory decision-making for medical devices. August 31, 2017. FDA-2016-D-2153
5. EMA. Journey towards a roadmap for regulatory guidance on real-world evidence. 19 February 2025. EMA/100019/2025
6. US Food and Drug Administration. Framework for FDA's Real-World Evidence Program. 2018 [cited 2025 Aug 25]. Available at: https://www.fda.gov/media/120060/download
7. McMurry TL, Hu Y, Blackstone EH, et al. Propensity scores: methods, considerations, and applications. J Thorac Cardiovasc Surg. 2015;150:14–19. doi:10.1016/j.jtcvs.2015.03.057

8. Brumback BA. Fundamentals of Causal Inference: With R (1st ed.). Chapman and Hall/CRC. 2021. doi:10.1201/9781003146674
9. Dang LE, Gruber S, Lee H, et al. A causal roadmap for generating high-quality real-world evidence. J Clin Transl Sci. 2023;7(1):e212, 1–12. doi:10.1017/cts.2023.635
10. Yoo-Jeong M, Waldrop-Valverde D, McCoy K, et al. A structural equation model of HIV-related symptoms, depressive symptoms, and medication adherence. HIV AIDS. 2016:2(3). doi:10.16966/2380-5536.123
11. Le-Rademacher J, Gunn H, Yao X, Schaid DJ. Clinical trials overview: from explanatory to pragmatic clinical trials. Mayo Clin Proc. 2023;98(8):1241–53. doi:10.1016/j.mayocp.2023.04.013
12. Edwards D, Best N, Crawford J, et al. Using Bayesian dynamic borrowing to maximize the use of existing data: a case-study. Ther Innov Regul Sci. 2024;58(1):1–10. doi:10.1007/s43441-023-00585-3
13. Emir B, Johnson K, Kuhn M, Parsons B. Predictive modeling of response to pregabalin for the treatment of neuropathic pain using 6-week observational data: a spectrum of modern analytics applications. Clin Ther. 2017;39(1):98–106. doi:10.1016/j.clinthera.2016.11.015
14. Wang J, Xu Y, Liu L., et al. Comparison of LASSO and random forest models for predicting the risk of premature coronary artery disease. BMC Med Inform Decis Mak. 23(1):297. doi:10.1186/s12911-023-02407-w
15. National Institutes of Health (US). Bethesda (MD): National Library of Medicine (US). Available from: https://clinicaltrials.gov
16. Verweij S, Weemers J, Jonker C, et al. Learning from European regulator-initiated studies for regulatory decision making. Drug Discov Today. 2025;30(1):104256. doi:10.1016/j.drudis.2024.104256
18. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol. 2008;61(4):344–9. doi:10.1016/j.jclinepi.2007.11.008
19. Benchimol EI, Smeeth L, Guttmann A, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. PLoS Med. 2015 Oct 6;12(10):e1001885. doi:10.1371/journal.pmed.1001885

Author information

Natasa Rajicic, ScD, has been a practicing biostatistician for more than 20 years. Her expertise includes adaptive study designs and implementation, oncology, and rare disease. She has served as an independent statistical reviewer as a member of monitoring boards. Earlier in her career, Natasa was a biostatistician at Arrowhead Pharmaceuticals, Cytel, Pfizer, Center for Biostatistics in AIDS Research in Boston, and taught Research Design at Columbia University in New York.

Anders Mattsson, PhD, has experience for more than 30 years in the field of biostatistics. Special focus has been on epidemiology and Phase III and Phase IV clinical trials. Special areas of interest have been on study design, statistical methods in epidemiology, ionising radiation epidemiology, pharmacoepidemiology, risk estimation and effectiveness analyses in observational studies. He has authored or co-authored over 70 peer-reviewed publications. He worked at the Karolinska Institute, for the regional cancer registry in Stockholm, and in the pharmaceutical industry before he retired in 2020.

Maria Kołtowska-Häggström, MD, PhD, runs Proper Medical Writing, the first Polish medical writing agency. She has 30 years of experience in the pharmaceutical industry, primarily in clinical research, including real-world data (RWD) and real-world evidence (RWE). Maria has authored over 80 peer-reviewed publications, many focused on RWE and patient-reported outcomes. She is a member of EMWA, EASE, ESE, and GRS; serves as a journal reviewer and Associate Editor for BMC Endocrine Disorders; and is an EMWA workshop leader and a section editor for *Medical Writing*.

Embracing the potential of real-world data: An industry perspective

Anna Woziwodzka[1],
Wendelgard Pisternick-Ruf[1],
Anouk Déruaz-Luyet[2]

[1] Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach an der Riss, Germany
[2] Global Integrated Evidence, Boehringer Ingelheim International GmbH, Ingelheim am Rhein, Germany

doi: 10.56012/rcwk9212

Correspondence to:
Anna Woziwodzka
anna.woziwodzka@boehringer-ingelheim.com

Abstract

Randomised controlled trials are the gold standard for evaluating the efficacy of medical interventions, offering robust evidence through controlled designs that minimise bias. However, their generalisability to diverse patient populations is typically limited. Real-world evidence (RWE), derived from real-world data, such as registries, electronic health records, claims databases, patient networks, social media, and wearables, has emerged as a vital complement to randomised controlled trials, addressing questions of effectiveness, safety, and cost-effectiveness across a medicinal product's lifecycle. In this article, we highlight the expanding role of RWE, from early development to post-launch activities, with examples from sponsor's RWE studies.

Image: Freepik

Introduction

For more than five decades, randomised controlled trials (RCTs) have been the gold standard for demonstrating the therapeutic benefit of medical interventions prior to marketing authorisation.[1] By employing standardised methods to minimise bias, such as randomisation and blinding, while comprehensively measuring outcomes to establish efficacy of a novel product, the controlled design of RCTs provides significant advantages for evidence generation. On the other hand, RCTs are typically resource-intensive, time-consuming, and often conducted in relatively homogenous patient populations defined by restrictive inclusion and exclusion criteria. Therefore, the results of RCTs may not always be generalisable to broader, more diverse patient populations, which can leave certain questions about effectiveness and long-term safety of an investigational product unanswered.[2]

Real-world data (RWD) are used by regulatory authorities for post-marketing safety assessment and surveillance of medicinal products, as well as by payers and health technology assessment bodies to inform cost-effective coverage decisions. RWD refers to data collected outside of randomised clinical trials and encompasses information on patient health status and the delivery of healthcare from routine sources, such as data from registries, electronic health records, insurance claims, patient networks, social media, and patient-generated data from wearables. Recent advancements in computing technologies and the widespread availability of electronic health data have allowed RWD play an increasing role in drug development and healthcare decision-making across diverse stakeholders.

Real world evidence (RWE), in turn, is defined as clinical evidence derived from the analysis of RWD, for example, to characterise a population and disease and to evaluate the utilization, benefits, and risks associated with medicinal products.[3,4] RWE plays a pivotal role in addressing diverse needs across the product's lifecycle, from early development to post-launch activities (Figure 1). Historically, RWE has been primarily used to assess disease statistics and fulfil

post-authorisation safety monitoring obligations, but it has more recently taken on a strategic role in all phases of the product lifecycle, from early development to post-launch.

RWE in early and late drug development

Optimal planning of RWE begins early, with a forward-looking approach to anticipate future needs. During early development, evidence generation focuses on supporting decisions related to the positioning of investigational products and informing trial design and start-up activities. This includes characterising patient populations, understanding disease burden, and gathering insights into clinical practice (Figure 1).

For example, the European Scleroderma Trials and Research (EUSTAR) registry has been used to generate RWE on systemic sclerosis-associated interstitial lung disease (SSc-ILD). A prospective analysis of EUSTAR registry data on changes in lung function collected over 5 years from 800 patients identified risk factors for SSc-ILD and patterns of progression.[6] Also, a retrospective analysis of 6,000 patients in the EUSTAR registry used a stepwise cohort enrichment approach to identify patients with SSc-ILD at risk of disease progression and determine how many would be eligible for trial inclusion, with the aim of informing the selection of inclusion criteria and target populations for clinical trials.[7]

Similarly, claims databases have been used to generate RWE on unmet needs in generalised pustular psoriasis (GPP). A study in Japan compared profiles of patients with GPP and plaque psoriasis with the general population regarding comorbidities, medication use, health care resource utilisation, and health care costs in a 1-year follow-up period. The study highlighted that GPP patients had a higher disease burden, greater reliance on systemic treatments, and increased healthcare utilisation than those with plaque psoriasis.[8] A further study of claims data in the US highlighted the significant economic burden of GPP and palmoplantar pustulosis, including higher healthcare costs and more frequent inpatient visits than in the general population.[9]

RWE post-launch (growth and mature phase)

Post-launch, RWE generation focuses on the long-term effectiveness and safety in routine clinical practice to support clinical trial findings and payer discussions (Figure 1), as well as clinical adoption of a new product. Also, post-authorisation safety studies (PASS) may be needed to fulfil post-marketing commitments to regulatory authorities. Such studies are often conducted in non-interventional settings to complement efficacy and safety data available at the time of initial marketing authorisation.

An example PASS study evaluated the risk of acute pancreatitis in patients with type 2 diabetes treated with metformin and initiating empagliflozin therapy.[10] The study was conducted to address emerging safety concerns suggesting a link between several glucose lowering therapies, including sodium-glucose cotransporter-2 inhibitors, and this relatively rare but potentially serious and occasionally fatal condition.[11,12] In the study, data from two large US claims databases were used to compare the incidence of acute pancreatitis between patients prescribed empagliflozin and those prescribed sulfonylurea. The results supported existing evidence that the use of empagliflozin in patients with type 2 diabetes is not associated with increased risk of acute pancreatitis.

RWE might also be needed to support

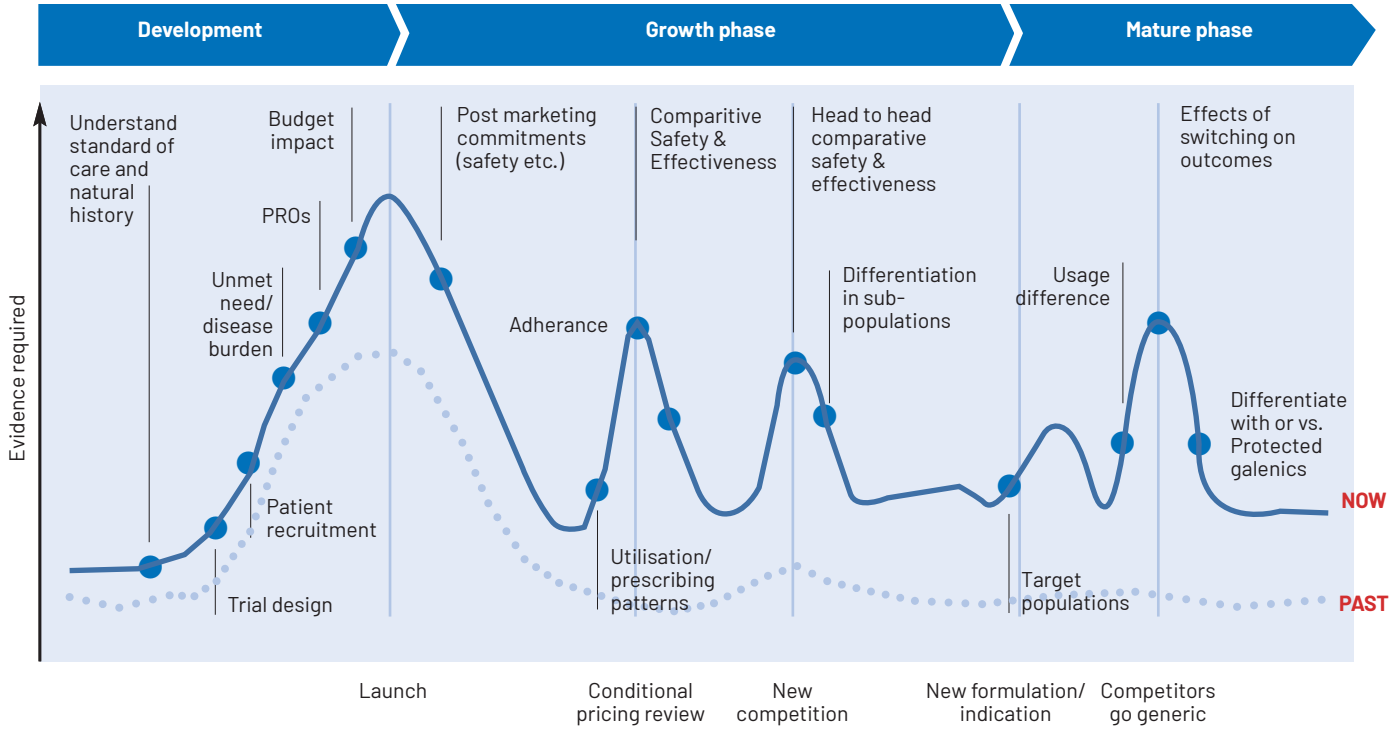


Figure 1. Real-world evidence generation along the lifecycle of a medicine. Adapted from Cerreta.[5]

Image: Freepik

different post-launch activities, such as pricing discussions, development of treatment guidelines, and clinical adoption.[13] In the mature phase of a medicine lifecycle, RWE helps address the impact of switching to generics on therapy outcomes and the preferences of patients (Figure 1).

The EMPRISE programme of studies: an example of RWE for bridging clinical trial findings with real-world practice

EMPA-REG OUTCOME, a prospective, placebo-controlled clinical trial, showed that patients with type 2 diabetes and established cardiovascular disease treated with empagliflozin had lower relative risks of cardiovascular death, all-cause mortality, and hospitalisation for heart failure.[15] Based on this trial, the US FDA expanded the indication for empagliflozin to reducing the risk of cardiovascular death in patients with type 2 diabetes and cardiovascular disease.[16] This was followed by respective changes in major clinical guidelines on diabetes treatment for patients with cardiovascular disease.[17,18]

Following the EMPA-REG OUTCOME trial, the EMPRISE (EMPagliflozin compaRative effectIveness and SafEty) program was initiated to assess the real-world impact of empagliflozin in patients with type 2 diabetes. This is a global, multi-year monitoring programme with a new-user, active-comparator cohort study design in which 1:1 propensity score-matching between patients initiating empagliflozin or a comparator was launched to evaluate the effectiveness, safety, and healthcare utilisation of empagliflozin in routine care across a broad spectrum of baseline cardiovascular risk.[14] Using data from the US, Europe, and Asia, the EMPRISE programme confirmed that, in a broader population than in the EMPA-REG OUTCOME trial, the risk of cardiovascular events, hospitalisations for heart failure, and mortality are lower in patients treated with empagliflozin than in those treated with a dipeptidyl peptidase-4 inhibitor (DPP-4i) or glucagon-like peptide-1 (GLP-1) receptor agonist.

Final analysis of the EMPRISE US study, which included over 115,000 matched pairs of patients initiating empagliflozin or a DPP-4i, further demonstrated that the risks of several cardiovascular outcomes, including myocardial infarction or stroke, hospitalisation for heart failure, major adverse cardiovascular events, and cardiovascular and all-cause mortality were lower in patients treated with empagliflozin.[19] The study also showed that the risk of diabetic ketoacidosis was higher and the risks of acute kidney injury, severe hypoglycaemia, and retinopathy progression were lower in patients initiating empagliflozin than in those initiating a DPP-4i.

Subgroup analysis in the EMPRISE US study further revealed benefits in subgroups of patients, such as older patients and those with a history of atherosclerotic cardiovascular disease or heart failure, adding to the evidence generated in the EMPA-REG OUTCOMES trial.[19] Finally, the EMPRISE Europe and Asia studies, which encompassed over 85,000 matched pairs of patients across 11 countries, validated the findings of the EMPRISE US study in international settings.[20] Thus, RWE generated by the EMPRISE programme complemented the results of RCTs by adding insights on the effectiveness of empagliflozin in patients with type 2 diabetes with or without of a history of cardiovascular disease or heart failure across diverse routine care models.

In addition to providing insight on the effectiveness of empagliflozin in patients with type 2 diabetes, EMPRISE enabled a direct comparison of the cardiovascular effects of empagliflozin and GLP-1 receptor antagonists that have demonstrated a cardioprotective

potential in clinical trials. EMPRISE showed that empagliflozin was associated with similar risks of myocardial infarction or stroke and lower risks of hospitalizations for heart failure and cardiovascular mortality than GLP-1 receptor antagonists.[21]

Furthermore, the EMPRISE programme also allowed cost-effectiveness and healthcare resource utilisation to be assessed. The EMPRISE US study showed that, in routine clinical practice, health care utilization and costs of care were lower in patients initiating empagliflozin than in those initiating a DPP-4i, irrespective of the underlying cardiovascular disease.[22] Analysis of EMPRISE data in Sweden also demonstrated that empagliflozin reduced the rates of hospitalisation and in- and outpatient visits in patients with type 2 diabetes.[23]

Real-world studies on tiotropium/olodaterol in chronic obstructive pulmonary disease: another example of RWE for bridging clinical trial findings with real-world practice

Real-world studies on tiotropium/olodaterol in patients with chronic obstructive pulmonary disease (COPD) also illustrate how RWE can be used post-launch. Using US healthcare claims and commercial laboratory data, Quint et al.[24] compared the real-world effectiveness and safety of two combination maintenance therapies for patients with COPD, tiotropium/olodaterol and long-acting β_2-agonists (LABA)/inhaled corticosteroids (ICS). The study showed that the risks of COPD exacerbations, pneumonia, and treatment escalation to triple therapy were lowered more by tiotropium/olodaterol than by LABA/ICS, highlighting the importance of tiotropium/olodaterol in avoiding ICS overuse and reducing the risk of exacerbations in patients with COPD.

Further post-launch RWE on tiotropium/olodaterol was generated in two studies evaluating treatment patterns in COPD using US and UK healthcare databases.[25,26] The studies showed that, despite existing guidelines recommending ICS only for patients with severe COPD meeting certain criteria, ICS are overprescribed in both the US and the UK, potentially putting patients at risk of side effects and increasing unnecessary healthcare costs.

Conclusion

RWE has become an essential component of evidence generation across the medicinal product lifecycle. RWE enhances understanding of patient populations, disease burden, and clinical practice, and it provides evidence to inform trial design and optimise positioning of medicines. Post-launch, RWE further encompasses evaluation of real-world effectiveness, safety, and cost-effectiveness, complementing findings from RCTs. By providing insights into cost-effectiveness and healthcare resource utilisation, RWE enhances understanding of a medicine's value, supporting discussions with payers and informing treatment planning from a pharmacoeconomic perspective.

Disclaimers

The opinions expressed in this article are the authors' own and not necessarily shared by their employer or EMWA.

Disclosures and conflicts of interest

AW and WP-R are employees of Boehringer Ingelheim Pharma GmbH & Co. KG. AD-L is an employee of Boehringer Ingelheim International GmbH.

References

1. Bothwell LE, Greene JA, Podolsky SH et al. Assessing the gold standard lessons from the history of RCTs. N Engl J Med. 2016;374(22):2175-81.
2. Sherman RE, Anderson SA, Dal Pan GJ et al. Real-world evidence – What is it and what can it tell us? N Engl J Med. 2016;375(23):2293-7. doi:10.1056/NEJMsb1609216
3. EMA. Update on Real World Evidence Data Collection. 2016 [cited 2025 May 20]. Available from: https://health.ec.europa.eu/system/files/2016-11/4_real_world_evidence_ema_presentation_0.pdf
4. FDA. Framework for FDA's real-world evidence. 2018 [cited 2025 May 20]. Available from: https://fda.report/media/120060/framework-FDAs-real-world-evidence-program_0.pdf
5. Cerreta F. EMA Adaptive Pathways Pilot Presentation [cited 2025 May 20]. Available from: https://health.ec.europa.eu/document/download/98a5c418-8019-437d-bc12-a3b0b6a0fb8e_en?filename=6_presentation_adaptive_pathways_learnings.pdf&prefLang=hu
6. Hoffmann-Vold AM, Allanore Y, Alves M et al. Progressive interstitial lung disease in patients with systemic sclerosis-associated interstitial lung disease in the EUSTAR database. Ann Rheum Dis. 2021;80:219-27. doi:10.1136/annrheumdis-2020-217455
7. Hoffmann-Vold A, Gahlemann M, Graf N et al. Feasibility and efficacy of cohort enrichment for progressive lung fibrosis in systemic sclerosis - a EUSTAR database analysis. ATS 2019, 115th Int Conf of the American Thoracic Society (ATS), Dallas, 2019-05-17; Am J Respir Crit Care Med. 2019;199;Abstr:A5614.
8. Okubo Y, Kotowsky N, Gao R et al. Clinical characteristics and health-care resource utilization in patients with generalized pustular psoriasis using real-world evidence from the Japanese Medical Data Center database. J Dermatol. 2021;48:1675-87. doi:10.1111/1346-8138.16084
9. Hanna ML, Singer D, Valdecantos WC. Economic burden of generalized pustular psoriasis and palmoplantar pustulosis in the United States. Curr Med Res Opin. 2021;37(5):735-42. doi:10.1080/03007995.2021.1894108
10. Fazeli Farsani S, Iglay K et al. Risk of acute pancreatitis among new users of empagliflozin compared to sulfonylureas in patients with type 2 diabetes: A post-authorization safety study. Pharmacoepidemiol Drug Saf. 2024;33(5):e5800. doi:10.1002/pds.5800
11. Azoulay L, Filion KB, Platt RW et al. Association between incretin-based drugs and the risk of acute pancreatitis. JAMA Intern Med. 2016;176(10):1464-73. doi:10.1001/jamainternmed.2016.1522
12. Palapra H, Viswam SK, Kalaiselvan V, et al. SGLT2 inhibitors associated pancreatitis: signal identification through disproportionality analysis of spontaneous reports and review of case reports. Int J Clin Pharmacol. 2022;44(6):1425-33. doi:10.1007/s11096-022-01476-7
13. Kent S, Meyer F, Pavel A, et al. Planning post-launch evidence generation: Lessons from France, England and Spain. Clin Pharmacol Ther. 2025;117(4):961-6. doi:10.1002/cpt.3586
14. Patorno E, Najafzadeh M, Pawar A, et al. The EMPagliflozin compaRative effectIveness and SafEty (EMPRISE) study programme: Design and exposure accrual for an evaluation of empagliflozin in routine clinical care. Endocrinol Diabetes Metab. 2019;3(1):e00103. doi:10.1002/edm2.103

15. Zinman B, Wanner C, Lachin JM, et al. Empagliflozin, cardiovascular outcomes, and mortality in type 2 diabetes. N Engl J Med. 2015;373(22):2117-28. doi:10.1056/NEJMoa1504720
16. FDA. Press announcement: FDA Approves Jardiance to Reduce Cardiovascular Death in Diabetics. 2016 [cited 2025 May 20]. Available from: https://wayback.archive-it.org/7993/20170111160646/http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm531517.htm
17. Davies MJ, D'Alessio DA, Fradkin J, et al. Management of hyperglycemia in type 2 diabetes, 2018. A consensus report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). Diabetes Care. 2018;41(12):2669-701. doi:10.2337/dci18-0033
18. American Diabetes Association. 9. Pharmacologic approaches to glycemic treatment: Standards of Medical Care in Diabetes–2021. Diabetes Care. 2021;44(suppl 1):S111-24. doi:10.2337/dc21-S009
19. Htoo PT, Tesfaye H, Schneeweiss S et al. Effectiveness and safety of empagliflozin: final results from the EMPRISE study. Diabetologia 2024;67(7):1328-42. doi:10.1007/s00125-024-06126-3
20. Vistisen D, Carstensen B, Patorno E, et al. Empagliflozin is associated with lower cardiovascular risk compared with dipeptidyl peptidase-4 inhibitors in adults with and without cardiovascular disease: EMPagliflozin compaRative effectIveness and SafEty (EMPRISE) study results from Europe and Asia. Cardiovasc Diabetol. 2023;22(1):233. doi:10.1186/s12933-023-01963-9
21. Htoo PT, Tesfaye H, Schneeweiss S et al. Cardiorenal effectiveness of empagliflozin vs. glucagon-like peptide-1 receptor agonists: final-year results from the EMPRISE study. Cardiovasc Diabetol. 2024;23(1):57.
22. Htoo PT, NajafZadeh M, Tesfaye H et al. Health care utilization and costs associated with Empagliflozin in older adults with Type 2 diabetes. Diabetes Care. 2024;47(11):1900-7. doi:10.2337/dc24-0270
23. Nyström T, Toresson Grip E, Gunnarsson J, et al. Empagliflozin reduces cardiorenal events, healthcare resource use and mortality in Sweden compared to dipeptidyl peptidase-4 inhibitors: real world evidence from the Nordic EMPRISE study. Diabetes Obes Metab. 2023;25(1):261-71. doi:10.1111/dom.14870
24. Quint JK, Montonen J, Esposito DB, et al. Effectiveness and safety of COPD maintenance therapy with Tiotropium/Olodaterol versus LABA/ICS in a US claims database. Adv Ther. 2021;38(5):2249-70. doi:10.1007/s12325-021-01646-5
25. Bloom CI, Montonen J, Jöns O et al. First Maintenance Therapy for Chronic Obstructive Pulmonary Disease: Retrospective Analyses of US and UK Healthcare Databases. Pulm Ther. 2022;8(1):57-74.
26. Bloom CI, Montonen J, Jöns O, et al. Treatment transitions in chronic obstructive pulmonary disease: Retrospective analyses of US and UK healthcare databases. Pulm Ther. 2022;8(1):75-93. doi:10.1007/s41030-021-00180-7

Author information

Anna Woziwodzka, PhD, has 8 years of medical writing experience and academic background in biophysics and clinical microbiology. Following her transition from academia, she worked as a freelance medical writer in scientific communications. Since 2021, she has been a regulatory medical writer at Boehringer Ingelheim, with a focus on supporting clinical studies and developing submission documents for inflammation indications.

Wendelgard Pisternick-Ruf, PhD, graduated as a pharmacist from the Tübingen University and the ETH Zurich. She has over 25 years of experience in the pharmaceutical industry, including pharmacovigilance, medical affairs, and clinical development, with more than 10 years as a regulatory medical writing professional at Boehringer Ingelheim, working on clinical documents related to trials and marketing authorisation applications across different therapeutic areas.

Anouk Déruaz-Luyet, PhD, MPH, has more than 15 years of experience in epidemiology and RWE across industry and academia, spanning across therapeutic areas and supporting diverse business functions. Previously, she conducted patient-centred clinical research in neurosciences, developing diagnostic tests, and neurorehabilitation programmes. She holds a dual academic background in Neurosciences and Public Health.

Don't miss!

The December 2025 edition

Safety Writing

As the regulatory landscape continues to evolve, the importance of precise and thorough safety reporting has never been more critical. This issue will provide insights into the latest methodologies, best practices, and innovative approaches that are shaping the future of safety writing. The issue will feature articles on the development and submission of safety data, offering expert guidance on handling complex safety data.

Guest Editors: Iva Cvetkovic and Pavle Simeunovic

The role of real-world evidence in post-market clinical follow-up

Laura C. Collada Ali[1],
Kelly Goodwin Burri[2],
Katharina Friedrich[3]

[1] Trilogy Writing & Consulting GmbH, Cogne, Italy
[2] Stryker, Selzach, Switzerland
[3] Katylistic GmbH, Basel, Switzerland

doi: 10.56012/eqqu1042

Correspondence to:
Laura C. Collada Ali
laura.collada@trilogywriting.com

Abstract

Post-market clinical follow-up (PMCF) is a mandatory, ongoing process under EU MDR 2017/745, aimed at ensuring the continued safety and performance of medical devices. This manuscript outlines the regulatory requirements, methodologies, and integration of real-world data (RWD) in PMCF activities. It highlights how manufacturers can use RWD from registries, retrospective studies, and user surveys among other sources to fill clinical evidence gaps and support regulatory compliance. Case studies illustrate practical applications of RWD in PMCF. A systematic and data-driven PMCF approach is essential for effective post-market surveillance and the protection of patient health.

Understanding the requirements of post-market clinical follow-up

Post-market clinical follow-up (PMCF) is an integral process of the European Union's Medical Device Regulation (MDR) 2017/745 to continually assess performance and safety once a medical device has entered the market.[1-3]

It is not a one-off activity but rather an ongoing process that occurs throughout the device's lifecycle, providing manufacturers with updated clinical evidence to support their device's conformity with regulatory requirements. This also includes the collection of clinical data from real-world use to further evaluate the device when it is used in a broader patient population.[2-3]

Regulatory framework and requirements

According to the MDR, manufacturers must establish and implement a post-market surveillance (PMS) system that includes PMCF as a crucial element.[1-3]

PMCF must be planned, systematic, and documented, outlining the objectives, methodology, and the clinical data to be collected. The collected data should then be analysed and used to update the clinical evaluation, risk management, PMS and other documents such as the summary of safety and clinical performance (SSCP), if applicable.[1-3]

PMCF methodologies

There are two primary types of PMCF: general and specific. General PMCF refers to the collection of clinical data that is not tied to a specific clinical question but is gathered as part of routine PMS activities.[2-3] This data may include, for instance, general feedback from healthcare professionals, information from systematic literature reviews, or data from vigilance databases.[2,3]

Specific PMCF, on the other hand, refers to targeted activities, such as high-quality user surveys, post-market studies or data collection from device registries. These activities are used to answer specific questions, e.g., from the clinical evaluation, and to close gaps in the clinical evidence of a medical device.[2,3]

Both general and specific PMCF activities must be well-documented and conducted in compliance with the MDR's requirements, and other international or local requirements (e.g., ISO 14155 or Good Clinical Practice) with clear objectives and methodologies. Manufacturers are also required to ensure that any clinical findings from PMCF are communicated to the relevant authorities, stakeholders, and users of the device.[1-3]

When real-world-data comes into play

Real-world-data (RWD) has become an invaluable resource in the post-market phase, particularly within the framework of PMCF. As healthcare evolves and patient care becomes more complex, RWD offers unique insights into the safety, effectiveness, and long-term performance of medical devices when used outside the controlled conditions of clinical investigations across diverse patient populations and clinical settings.

Unlike prospective interventional or randomised controlled trials (RCTs), RWD

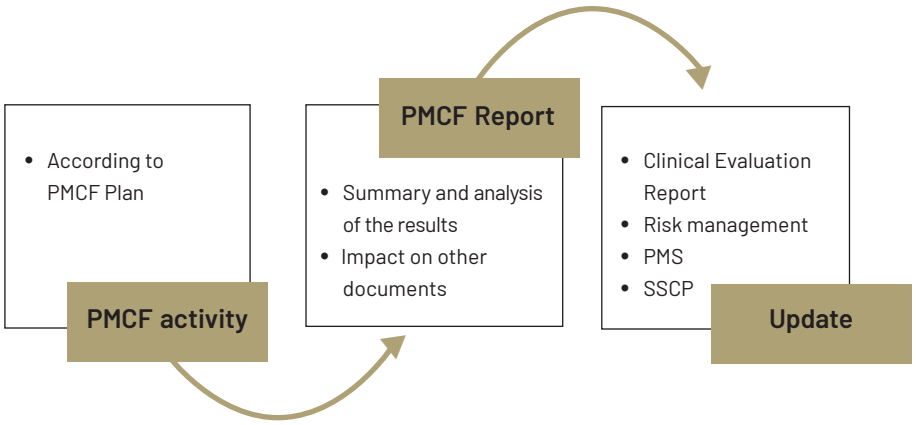


Figure 1. PMCF close connections with other processes

Abbreviations: PMCF, post-market clinical followup; PMS, post-market surveillance; SSCP, Summary of Safety and Clinical Performance.

Table 1. Comparison of data from standard clinical investigations and RWE

Data from "standard" clinical investigations	Data from RWE
Carefully selected inclusion and exclusion criteria	Not limited to a selected patient group
Intended to address a certain hypothesis or clinical question	To confirm the performance and safety in a routine clinical setting
Limited to a predefined sample size and time period	Not necessarily limited to a sample size and suitable to collect data over the device lifetime

reflects the broader population, including patients with various comorbidities, varying degrees of disease severity, and other factors that may not be well represented in traditional clinical trials. The use of RWD provides a more accurate and comprehensive understanding of how devices perform in real-world settings, helping to identify issues that may not have been detected during pre-market evaluation.[4] (See Table 1).

Types of real-world data

There are numerous sources of RWD that can be leveraged to support PMCF activities.[5,6] Some of the most widely used types of RWD are shown in Table 2.

Table 2. Overview of data sources for RWE

Electronic health records (EHRs)	**Retrospective chart review**
EHRs provide large real-world data sets with information about diagnoses, treatment plans and results, and might also include medical device brand names, catalogue numbers/UDI. We also use this data in combination with insurance claims databases. More data vendors are also providing services to analyse unstructured data (clinical notes, imaging) to enable more sophisticated analysis of device performance outcomes. Manufacturers should be aware of potentially limited data quality and bias due to the design or analysis of EHRs.	Retrospective reviews of medical records are a great source of real-world performance and safety data that don't rely on user compliance as surveys. Although they require a predefined study protocol and statistical analysis plan, they are more cost-effective than prospective studies with a lower selection bias.
Registries	**Laboratory information systems (LIS)**
Registries are great sources of RWE with the potential to collect long-term data. So far, national registries are limited to certain devices, such as orthopaedic implants. As an alternative, manufacturers can initiate their own medical device registries. Manufacturer-initiated registries usually follow a predefined study protocol that can be targeted to the collection of the most relevant performance and safety parameters.	Laboratory information systems store and manage data related to laboratory tests, including blood work, diagnostic imaging, and other tests that inform patient care. These systems can provide important real-world data on how devices perform in relation to specific diagnostic or therapeutic procedures. For instance, devices such as in-vitro diagnostic (IVD) tests or point-of-care testing devices can be tracked through LIS to assess their accuracy, reliability, and potential for misdiagnosis in clinical practice.
Surveys	**Social media listening**
Surveys can be used to collect data from healthcare professionals or patients. They can be designed to collect general feedback on user experience and user satisfaction or to collect data on specific procedures and device usages. Despite their known limitations (e.g., low response rates), they are great tools to reach various users in a relatively short time.	Social media listening identifies early signals of safety issues, user concerns, and real-world device performance in general. By monitoring public posts, reviews, and discussions, manufacturers can detect adverse events, misuse, or unmet needs not captured through traditional channels. Analysing this user-generated content provides timely insights; still, it can be difficult to weight against other sources of data which are monitored by health-care professionals.

Image: Freepik

The Role of RWD and RWE in PMCF

The integration of RWD into PMCF activities allows for the continuous monitoring of medical devices' safety and performance.[3] By analysing data from a range of real-world sources, manufacturers can identify emerging risks, assess long-term performance, and make necessary adjustments to their products or PMS and PMCF plans.[1–4,6] RWD provides regulators with a more comprehensive understanding of a device's performance across diverse patient populations and clinical settings.

They can also help bridge gaps in clinical investigation data, particularly for devices that are used in rare conditions with unique patient populations, ensuring that regulatory decisions are based on the best available evidence.[6] As an example, we typically use RWD to support very narrow indications (e.g., distal femur fracture with intra-articular extension). Another important use is to provide paediatric data. In both cases, running a traditional clinical investigation to collect this data would be very time consuming (long enrolment with few sites, hurdles for approvals of paediatric studies) and expensive.

Case study 1 – Registries for RWE

Situation: A manufacturer of orthopaedic implants, which are considered as medium to high-risk devices, wanted to use publicly available joint prosthesis registries to retrieve performance and safety data for their medical device. National joint registries are valuable sources of RWE, especially for orthopaedic implants, due to their long-term follow-up data.[6]

Problem: However, these registries typically do not provide device-specific data in their standard annual reports, limiting manufacturers' ability to assess and compare individual device performance.

Solution: To address this, an orthopaedic implant manufacturer requested two additional device-specific reports from the registry owner: one focusing on their own device and another on a group of benchmark devices. These reports enabled direct comparison with the State of the Art and will now be received annually. This approach enhances the value of the registry as a continuous RWE source.

Potential challenge: Smaller registries may not have the resources available to generate custom reports for manufacturers or may not agree to provide data for comparator products.

Case study 2 – Retrospective medical records review for RWE

Situation: A manufacturer of vascular stents, considered as high-risk devices, had strong clinical evidence supporting the device's use in lower leg arteries, aligning with part of its intended use.

Problem: The device's broad indication – including use in upper thigh arteries – lacked robust clinical evidence, relying only on isolated case reports.

Solution: To address this gap, the manufacturer identified a hospital that frequently used the stent for upper thigh lesions. They conducted a retrospective analysis of the hospital's database, successfully gathering performance and safety data to support the broader indication.

Potential challenge: This is not typically a continuous activity, and a single centre may not have sufficient volume to provide enough cases for the specific indication.

Case study 3 – User surveys for RWE

Situation: A manufacturer offered low risk medical devices primarily used as accessories in interventional procedures. These devices had a low-risk profile and were not typically featured in scientific publications.

Problem: Given their accessory role and simplicity, it was neither practical nor necessary to conduct clinical studies, yet the manufacturer still needed performance and safety data to support the device's use.

Solution: The manufacturer implemented a user survey using a simple case report form to be completed during or immediately after an intervention with the device. This approach enabled the collection of relevant data on technical performance and potential safety events. Short-term follow-up was sufficient due to the device's nature and intended use.

Potential challenge: User surveys for RWE in PMCF may face challenges with response bias, limited clinical depth, and inconsistent data quality, potentially undermining the reliability of safety and performance insights.

Case study 4 – Social media listening for RWE

Situation: A manufacturer of wearable cardiac monitors aimed to enhance post-market surveillance by exploring non-traditional data sources.

Problem: Despite formal reporting channels, some users shared device issues –such as skin irritation or inaccurate readings – only through social media platforms. These signals were missed by conventional PMS systems.

Solution: The manufacturer implemented a social media listening tool to monitor public posts related to their product. This enabled early identification of recurring user complaints, prompting further analysis of the published literature, complaints and incidents databases, and other sources. The approach improved patient safety and supplemented traditional PMS data.

Potential challenge: It may be more cost-effective for companies to embed social media listening as part of an overall vigilance strategy rather than using for a single product PMCF needs.

Conclusion

Post-market clinical follow-up (PMCF) is a critical component of the EU MDR framework, ensuring that medical devices continue to meet safety and performance standards throughout their lifecycle.[1] As demonstrated in this manuscript, PMCF must be systematic, targeted, and responsive to evolving clinical needs and regulatory expectations. The integration of RWD significantly enhances the PMCF process by providing timely, relevant insights from diverse patient populations and clinical settings. Whether through registries, retrospective studies, or user surveys, leveraging RWD enables manufacturers to close evidence gaps, refine risk management, and maintain regulatory compliance. As healthcare systems and data infrastructures evolve, robust PMCF strategies grounded in real-world evidence will be essential for ensuring device safety and public health.

Acknowledgements

The authors would like to thank Anuradha Alahari for reviewing this article and providing valuable support.

Disclosures and conflicts of interest

The author declares no conflicts of interest.

References

1. European Parliament and the Council of the European Union. Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices. Off J Eur Union. 2017;L117:1–175. Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32017R0745
2. International Organization for Standardization. ISO 14155:2020: Clinical investigation of medical devices for human subjects – Good clinical practice. Geneva: ISO; 2020.
3. Medical Device Coordination Group. MDCG 2020-7: Guidance on PMCF plan template. 2020. Available from: https://health.ec.europa.eu/system/files/2020-09/md_mdcg_2020_7_guidance_pmcf_plan_template_en_0.pdf
4. Medical Device Coordination Group. MDCG 2020-8: Guidance on PMCF evaluation report template. 2020. Available from: https://health.ec.europa.eu/system/files/2020-09/md_mdcg_2020_8_guidance_pmcf_evaluation_report_en_0.pdf
5. Makady A, de Boer A, Hillege H, et al. What is real-world data? A review of definitions based on literature and stakeholder interviews. Value Health. 2017;20(7):858-65. Available from: doi.10.1016/j.jval.2017.03.008
6. Collada Ali LC, Friedrich K, Pritchard G. Post-market clinical follow-up insights. Med Writ. 2022;31(2):67–73. Available from: https://journal.emwa.org/medical-devices/post-market-clinical-follow-up-insights/

Author information

Laura C. Collada Ali s a Senior Medical Writing Manager at Trilogy Writing & Consulting. She has over 25 years of experience writing, editing, and managing medical and scientific documentation for diverse clients and audiences worldwide. She is a member of EMWA's Executive Committee and a workshop leader and holds multiple certificates in linguistics, pharmacology, medical devices, clinical research, and regulatory writing.

Kelly Goodwin Burri has more than 20 years of experience in medical writing, clinical research, and epidemiology. She is currently a Manager of Real-World Evidence at Stryker where she is leveraging RWE to meet clinical evaluation and post-market requirements for medical devices.

Katharina Friedrich is a medical writer with experience in MDR regulatory writing and is the founder of Katylistic GmbH in Basel, Switzerland. She prepares Clinical Evaluation Plans and Reports, PMCF Plans and Reports and SSCPs in compliance with MDR 2017/745 for class I to class III devices.

From algorithms to insights: The role of medical writers in AI-enhanced real-world evidence

Gomathi Priya Jeyapal, Prarthana Vigneshwari Reddy, Vrushabh Baburao Satav, Pattabhi Machiraju
Indegene, Bengaluru, Karnataka, India

doi: 10.56012/vysp1464

Correspondence to:
Gomathi Priya Jeyapal
priya.jeyapal@indegene.com

Abstract

Real-world evidence (RWE) complements randomised controlled trials (RCTs) by assessing treatment effectiveness in diverse populations. Integrating artificial intelligence (AI) and machine learning (ML) enhances RWE by enabling predictive modelling, risk stratification, and clinical decision support. ML techniques like supervised or unsupervised learning, logistic regression, decision trees, random forests, and XGBoost can help optimise regulatory decision-making and patient care. This paper explores how the AI/ML models help identify high-risk patients, predict disease progression, and assess healthcare burden. The medical writer's role in structuring findings into clinically meaningful insights is essential for bridging the gap between data science and clinical application. As AI advances, skilled medical writers will ensure transparency, ethical compliance, and effective communication of AI-driven RWE findings.

Introduction

Randomised controlled trials (RCTs) are the gold standard for establishing causality in controlled settings, yet their findings often lack applicability to diverse real-world populations with varying genetic backgrounds, comorbidities, and treatment regimens. This "efficacy-effectiveness gap" limits the generalisability of RCT outcomes, posing challenges for regulatory decision-making.[1] To bridge this gap, healthcare stakeholders, including pharmaceutical companies, regulatory agencies, and health technology assessment (HTA) organizations, increasingly integrate real-world data (RWD) with RCTs.[2] Enabled by technological advancements, RWD comprising electronic health records (EHRs), registries, claims data, and mobile health applications offers valuable insights into routine healthcare delivery and patient outcomes.

When analysed, RWD generates real-world evidence (RWE), informing treatment effectiveness, safety, and economic impact, thereby supporting data-driven healthcare decisions and enhancing clinical and regulatory strategies.[3,4] The healthcare industry is witnessing an unprecedented transformation driven by artificial intelligence (AI) and machine learning (ML), which are redefining RWE generation, interpretation, and utilisation.[5] RWD, encompassing EHRs, claims data, patient registries, and wearable device outputs, offers vast potential to complement traditional clinical trials.[6]

AI identifies the patterns, medical writers connect them to patients, policies, and practice.

AI/ML algorithms enable efficient processing, analysis, and predictive modelling of these complex datasets, leading to actionable insights that improve patient outcomes, optimise treatment pathways, and guide regulatory decisions.[7] AI is significantly transforming the use of RWE in healthcare by facilitating more precise data analysis and decision-making. AI technologies are instrumental in analysing vast and complex RWD sources. These AI-driven approaches help identify patterns in treatment responses, predict patient outcomes, and improve clinical decision-making by integrating RWD into the healthcare system.[8,9]

However, as AI continues to revolutionize healthcare research, a critical challenge has emerged: the communication of complex, algorithm-driven insights to various healthcare stakeholders.[10] Regulators, clinicians, pharmaceutical companies, and policymakers require clear, precise, and scientifically accurate interpretations of AI-generated RWE.[11] Medical writers serve as essential intermediaries, ensuring that interpretations of AI-generated RWE are not only methodologically sound but also comprehensible, regulatory-compliant, and aligned with healthcare decision-making frameworks.

This paper explores the evolving landscape of AI-driven RWE, highlighting key ML techniques, such as clustering techniques, dimensional reduction algorithms, logistic regression (LR), decision trees (DT), random forests (RF), and extreme gradient boosting (XGBoost). It also delves into the role of medical writers in bridging the gap between complex AI/ML outputs and stakeholder needs, ensuring that scientific narratives derived from AI/ML-driven analytics are both accessible and impactful.

AI/ML techniques in RWE generation

The integration of AI/ML into RWE research enables the extraction of valuable patterns and insights from vast datasets. Supervised, unsupervised, and reinforcement learning are core ML approaches applied in healthcare and RWE generation. Supervised learning uses labelled data with known outcomes to train predictive models – such as LR, DT, RF, and XGBoost – that are commonly applied in diagnosing conditions like diabetes or hypertension. This approach follows a defined, iterative process involving data selection, processing, model training, and evaluation using metrics like receiver operating characteristic (ROC) curves and confusion matrices.

Unsupervised learning, on the other hand, works with unlabelled data to identify hidden structures or patterns using techniques such as clustering algorithms (e.g., k-means, hierarchical clustering) and dimensionality reduction methods like principal component analysis (PCA), which help group patients or detect

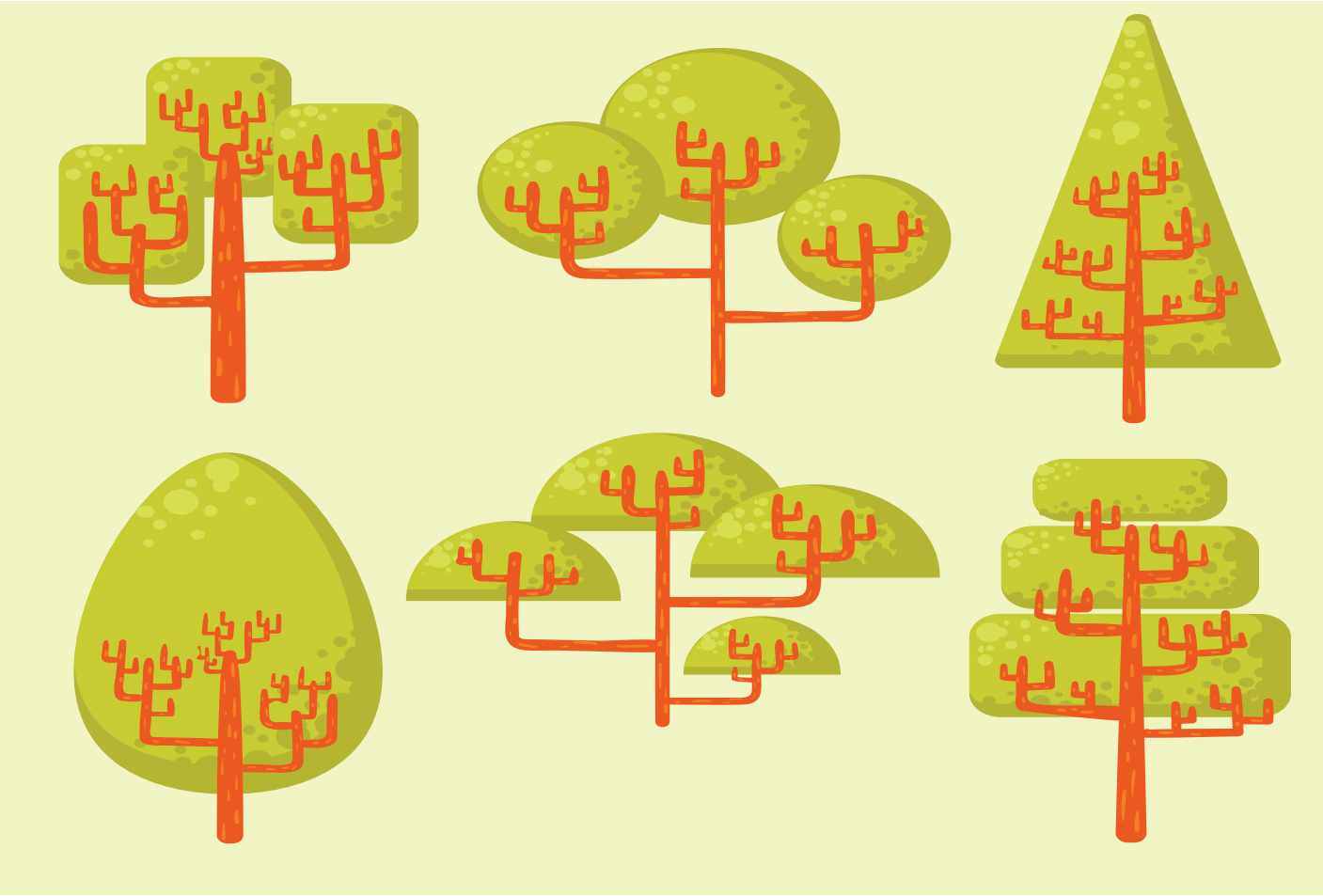


Image: Freepik

anomalies without predefined outcomes, though careful interpretation is necessary. Reinforcement learning allows systems to learn optimal decisions through trial and error, guided by feedback or rewards, making it promising for dynamic treatment decision support, despite challenges in defining rewards and causal pathways. The following section highlights key ML techniques and explores their applications in predicting clinical outcomes.

Supervised learning approaches

Logistic regression (LR)

LR is a foundational ML technique widely used in healthcare for binary classification tasks. It models the probability of an event occurring as a function of predictor variables, making it valuable for predicting clinical outcomes, adverse events, and patient risk stratification.[12] LR is a valuable tool for identifying high-risk patient groups by calculating a probability score that classifies patients into high-risk or low-risk categories, helping prioritise those needing immediate care. LR models are particularly useful in healthcare settings where predicting patient outcomes based on historical data can significantly enhance patient safety and clinical decision-making.[13] Performance metrics such as accuracy, precision, recall, F1-score (harmonic mean of precision and recall), and area under the curve (AUC)-ROC assess how well the model classifies patients.[14]

The model's coefficients show the contribution of each factor to a patient's risk level, with positive coefficients indicating a higher likelihood of being high-risk. The confusion matrix and ROC curve offer insights into how well the model distinguishes between high- and low-risk patients or slow and fast progressors. By analysing these results, healthcare providers can identify key risk factors and apply targeted interventions for better patient outcomes.[15]

> As real-world evidence evolves, medical writing expertise ensures that data-driven insights are ethically sound and clinically actionable.

Decision trees (DTs)

DTs use hierarchical structures to segment patient populations based on predictor variables, making them effective for classification and regression problems. A DT starts with a root node representing the entire dataset, which is then split into branches based on feature values.[16] These branches lead to decision nodes and eventually to leaf nodes, which represent the outcome. For example, a DT model was used to predict the risk of cardiovascular events in a large patient data set by analysing factors such as blood pressure and cholesterol levels, providing a clear pathway for targeted interventions.[17] Each decision helps categorise a patient as high-risk or low-risk. The model is easy to interpret, as it shows exactly how the decision is made at each branch.[18]

A DT approach may be better than LR due to its ability to capture non-linear relationships and handle mixed data types. While LR assumes a linear relationship between the predictors and the outcome, DTs can model complex interactions between variables without requiring

Image: Freepik

any assumptions about the data distribution. This flexibility allows DTs to provide more accurate predictions in cases where the relationships between variables are not straightforward.[18] One of the key strengths of DTs is their interpretability. The model visually represents how decisions are made at each branch, showing the exact conditions that lead to a particular classification. This transparency is particularly valuable in healthcare, where understanding the rationale behind risk predictions is crucial for clinical decision-making.

Healthcare providers can easily follow the tree structure to see how different patient-related factors contribute to the risk classification. DTs can also predict the importance of features, highlighting which variables have the most significant impact on the model's predictions. Through understanding the importance of different features, healthcare providers can focus on the most influential factors when designing interventions to reduce readmission rates. Healthcare providers can also gain deeper insights into patient-related factors and their impact on risk classification by leveraging DTs. This clarity helps in identifying patients who are likely to transition to higher risk categories, enabling targeted interventions that improve patient outcomes.[18]

Random forests (RFs)

RF is a powerful ensemble learning method that leverages multiple DTs to improve the accuracy and robustness of predictions. Unlike a single DT, which can easily overfit and perform poorly on unseen data, an RF builds many DTs using random sampling (bootstrap sampling) of both the data points and the features at each split.[19] This aggregation of multiple trees ensures that individual tree biases and variance are minimised, resulting in a more stable and reliable model.[20] An RF can handle complex interactions between features and make better predictions, especially when simpler models like LR or single DT do not yield satisfactory results.[21]

Each tree in an RF is built from a random subset of the data and features, which helps to reduce overfitting and improve generalization to new data. The final prediction is made by averaging the predictions of all the trees in the forest, which enhances the model's accuracy and robustness. For example, an RF model was used to predict phenotype transformations by analysing complex genetic and environmental interactions. This approach improved the accuracy of predictions and helped identify key factors driving phenotype changes, providing valuable insights for personalised medicine.[22] RF models can also provide insights into feature importance, indicating which variables have the most significant impact on the predictions.

XGBoost

XGBoost is a highly efficient and powerful boosting algorithm that builds trees sequentially, with each tree correcting the errors of the previous one by focusing on misclassified data points, thereby improving the model's predictive power and accuracy.[19] It optimises both speed and performance by using regularisation techniques to prevent overfitting and by implementing efficient tree-building algorithms.[23] XGBoost has several advantages compared to other algorithms, such as its ability to handle missing data and highly parallelizable code in large and complex datasets. It employs a novel sparsity-aware algorithm for sparse data and a weighted quantile sketch for approximate tree learning. This makes it particularly suited for applications in healthcare, where datasets can be vast and intricate.

In predicting patient risk for transitioning to an advanced disease stage, XGBoost can analyse a multitude of variables, including genetic factors, medical history, and lifestyle habits. By identifying small patterns and interactions that may be crucial for accurate predictions, XGBoost provides a robust tool for risk stratification. For instance, in a study predicting the progression of chronic kidney disease, XGBoost outperformed other models by accurately identifying patients at high risk of rapid disease progression.[24] XGBoost also provides insights into feature importance, highlighting which variables have

the most significant impact on the model's predictions. This information is valuable for healthcare providers as it helps identify key risk factors and prioritise interventions.

Unsupervised learning algorithms

Unsupervised learning algorithms are used in RWD analysis to detect clusters, reduce dimensionality, and uncover latent structures. They are increasingly featured in health economics and outcomes research (HEOR), pharmacovigilance, and post-market surveillance. Common techniques used are clustering (e.g., k-means, hierarchical clustering), which groups similar patients based on multiple clinical and demographic variables, and dimensionality reduction (e.g., PCA, t-SNE), which simplifies high-dimensional data to reveal visual patterns or key contributors.

The key outputs are cluster assignments, group labels, visualisations (heatmaps, dendrograms, scatter plots), and metrics (silhouette score, variance explained, cluster centroids). In one example, unsupervised clustering helped identify three distinct patient subtypes within the chronic kidney disease population.[25] One subgroup, comprising 30% of the population, showed frequent treatment switching and higher hospitalisation rates, indicating a high-risk phenotype with possible unmet needs.

The contribution of medical writers

Medical writing encompasses a broad and complex field, from clinical trials to regulatory submissions and from medical education to patient communication. Medical writers play a crucial role in translating complex medical information from research studies, clinical trials, and scientific articles into clear content. Balancing scientific accuracy using reliable evidence with clarity for the intended audience is a key challenge in healthcare. By adhering to ethical standards, medical writers maintain the trust of the scientific community and public while enhancing medical practices and knowledge.[26] Table 1 shows the key responsibilities of medical writers in translating RWE model results.[27]

Medical writing best practices in RWE studies

In the context of RWE generation, medical writers serve as critical knowledge translators responsible for accurately interpreting, contextualising, and communicating complex data to a wide array of stakeholders. Several case studies and whitepapers highlight the role of medical writers in successful AI-driven RWE projects.[28,29] For example, a study using AI to analyse EHRs for predicting cardiovascular outcomes required medical writers to translate complex ML models into actionable insights for clinicians.[30] Another case involved the use of natural language processing to extract RWE from unstructured clinical notes,[31] with medical writers ensuring the findings were accurately represented in regulatory submissions.

As well as RWE generation, AI/ML have been used for generating medical text. Despite advances in text generation, AI/ML cannot replace human medical writers and their use in medical writing raises ethical concerns.[32] AI-generated text has the potential to perpetuate bias, misinformation, and plagiarism. Furthermore, computer models need to be retrained regularly to ensure they are up to date, as the field of medicine is constantly evolving. Given these concerns, medical writers are indispensable for safeguarding the integrity of medical information and its compliance with ethical and regulatory standards.

By breaking down complex data, working with different teams, ensuring transparency, and getting results ready for publication, medical writers help make AI insights easier to under-

Table 1. Key responsibilities of medical writers in translating results of real-world evidence models

Key responsibility	Description
Bridging the language gap	Medical writers act as translators between data scientists and healthcare audiences. They interpret model assumptions and methodology, data inputs and limitations, and outputs, such as risk scores, clusters, or probabilities.
Distilling key insights	Writers identify and emphasise results that reveal clinically significant subgroups, suggest treatment response differences, and indicate burden of disease or health outcomes.
Crafting clinically meaningful narratives	An example: "The model identified a patient segment at threefold higher risk of hospitalization within 12 months, characterised by polypharmacy, diabetes, and advanced age. This suggests a target for early intervention." This kind of narrative brings model results into the clinical and strategic realm.
Visual interpretation support	Writers also help develop or adapt visualisations (e.g., Kaplan-Meier curves, cluster heatmaps), annotate plots to highlight clinically relevant findings, and ensure visual outputs are publication – or submission-ready.
Ensuring scientific and regulatory rigour	In RWE deliverables, especially for HTA or regulatory use, writers must clearly describe the model type, inputs, and limitations, avoid over-interpretation of exploratory analyses, and frame findings within accepted scientific standards.
Communicating data generation techniques using AI/ML	Developing peer-reviewed manuscripts and conference presentations that highlight the value of AI/ML techniques in RWE generation.[27]

Abbreviations: AI, artificial intelligence; HTA, health technology assessment; ML, machine learning; RWE, real-world evidence.

stand and use in healthcare. As such, medical writers are essential for clearly and accurately communicating complex scientific information and making sure that AI-derived findings meet ethical and regulatory standards. Since AI continues to shape healthcare, the need for skilled medical writers will grow.

Conclusion

Medical writers are indispensable in the AI-enhanced RWE ecosystem, ensuring that complex data is transformed into meaningful insights. By collaborating with data scientists, clinicians, and regulators, medical writers can help unlock the full potential of AI in healthcare. Medical writers do not need to be data scientists, but they must understand the fundamentals of analytic methodologies. As the field evolves, medical writers must embrace AI as a tool for innovation, while maintaining the highest standards of scientific integrity and ethical communication.

Acknowledgements

Not applicable.

Disclaimers

The opinions expressed in this article are the authors' own and do not necessarily reflect the position of the authors' affiliated organisations.

Disclosures and conflicts of interest

The authors declare no conflicts of interest.

References

1. Makady A, de Boer A, Hillege H,et al.; GetReal Consortium. What is real-world data? A review of definitions based on literature and stakeholder interviews. Value Health. 2017;20(7):858-65. doi:10.1016/j.jval.2017.03.008
2. Pongiglione B, Torbica A, Blommestein H, et al. Do existing real-world data sources generate suitable evidence for the HTA of medical devices in Europe? Mapping and critical appraisal. Int J Technol Assess Health Care. 2021;37(1):e62. doi:10.1017/S0266462321000301
3. McNair D, Lumpkin M, Kern S, et al. Use of RWE to Inform regulatory, public health policy, and intervention priorities for the developing world. Clin Pharmacol Ther. 2022;111(1):44-51. doi:10.1002/cpt.2449
4. Khosla S, White R, Medina J, et al. Real world evidence (RWE) – a disruptive innovation or the quiet evolution of medical evidence generation? F1000Res. 2018;7:111. doi:10.12688/f1000research.13585.2
5. Olczak J, Pavlopoulos J, Prijs J, et al. Presenting artificial intelligence, deep learning, and machine learning studies to clinicians and healthcare stakeholders: an introductory reference with a guideline and a Clinical AI Research (CAIR) checklist proposal. Acta Orthop. 2021;92(5):513-25. doi:10.1080/17453674.2021.1918389
6. Kandhare P, Kurlekar M, Deshpande T, et al. A review on revolutionizing healthcare technologies with AI and ML Applications in pharmaceutical sciences. Drugs Drug Candidates. doi:10.3390/ddc4010009
7. Kolluri S, Lin J, Liu R,et al. . Machine learning and artificial intelligence in pharmaceutical research and development: a review. AAPS J. 2022;24(1):19. doi:10.1208/s12248-021-00644-3
8. Wynants L, van Calster B, Collins GS, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. BMJ. 2020;369:m1328. doi:10.1136/bmj.m1328
9. Knevel R, Liao KP. From real-world electronic health record data to real-world results using artificial intelligence. Ann Rheum Dis. 2023;82(3):306-11. doi:10.1136/ard-2022-222626
10. Alanazi A. Using machine learning for healthcare challenges and opportunities. Inform Med Unlocked. 2022;30:100924. doi:10.1016/j.imu.2022.100924
11. Dang A. Real-world evidence: A primer. Pharmaceut Med. 2023;37(1):25-36. doi:10.1007/s40290-022-00456-6
12. Shipe ME, Deppen SA, Farjah F, et al. Developing prediction models for clinical use using logistic regression: an overview. J Thorac Dis. 2019;11(Suppl 4):S574-84. doi:10.21037/jtd.2019.01.25
13. Coz E, Fauvernier M, Maucort-Boulch D. An overview of regression models for adverse events =analysis. Drug Saf. 2024;47(3):205-16. doi:10.1007/s40264-023-01380-7
14. Manikandan G, Pragadeesh B, Manojkumar V, et al. Classification models combined with Boruta feature selection for heart disease prediction.Inform Med Unlocked. . 2024;44:101442. doi:10.1016/j.imu.2023.101442.=
15. Maksabedian Hernandez EJ, Tingzon I, Ampil L, et al. Identifying chronic disease patients using predictive algorithms in pharmacy administrative claims: an application in rheumatoid arthritis. J Med Econ. 2021;24(1):1272-9. doi:10.1080/13696998.2021.1999132
16. Saha D, Manickavasagan A. Machine learning techniques for analysis of hyperspectral images to determine quality of food products: A review. Curr Res Food Sci. 2021;4:28-44. doi:10.1016/j.crfs.2021.01.002
17. Díaz JJS. Chapter 13 – Artificial intelligence in cardiovascular medicine: Applications in the diagnosis of infarction and prognosis of heart failure. In: Barh D, editor. Artificial Intelligence in Precision Health. Cambridge, MA: Academic Press; 2020:313-28. doi:10.1016/B978-0-12-817133-2.00013-6
18. Ozcan M, Peker S. A classification and regression tree algorithm for heart disease modeling and prediction. Healthc Anal (NY). 2023;3:100130. doi:10.1016/j.health.2022.100130
19. Mahajan P, Uddin S, Hajati F, et al. A comparative evaluation of machine learning ensemble approaches for disease prediction using multiple datasets. Health Technol. 2024;14(3):597-613. doi:10.1007/s12553-024-00835-w
20. Mienye ID, Sun Y, Wang Z. An improved ensemble learning approach for the prediction of heart disease risk. Inform Med Unlocked. 2020;20:100402. doi:10.1016/j.imu.2020.100402
21. Breiman L. Random forests. Machine learning. 2001;45(1):5-32. doi:10.1023/A:1010933404324
22. Matsuo R, Yamazaki T, Suzuki M, et al. A random forest algorithm-based approach to capture latent decision variables and their cutoff values. J Biomed Inform. 2020;110:103548 doi:10.1016/j.jbi.2020.103548

23. Hassoon IM. Boosting Learning Algorithms for Chronic Diseases Prediction: A Review. Iraqi J Computers Informatics. 2024;50(2):22-30. doi:10.25195/ijci.v50i2.506
24. Wu J, Gao Q, Tian M, et al. Explainable machine learning prediction of 1-year kidney function progression among patients with type 2 diabetes mellitus and chronic kidney disease: a retrospective study. QJM. 2025 Apr 24:hcaf101. Epub ahead of print. doi:10.1093/qjmed/hcaf101
25. Dashtban A, Mizani MA, Pasea L, et al. Identifying subtypes of chronic kidney disease with machine learning: development, internal validation and prognostic validation using linked electronic health records in 350,067 individuals. EBioMedicine, 2023;8. doi:10.1016/j.ebiom.2023.104489.
26. Khan DI. Unveiling the Vital Contributions of Professional Medical Writers in the Evolving Healthcare Landscape. Indian Pract. 2023;76(12):24-26.
27. Bergstrand S, Heddle C, Sabaté M, et al. Embracing artificial intelligence in medical writing: A new era of efficiency and collaboration. Med Writ. 2023;32(3):82-7. doi:10.56012/iamc1709
28. Franklin JM, Liaw KL, Iyasu S, et al. Real-world evidence to support regulatory decision making: New or expanded medical product indications. Pharmacoepidemiol Drug Saf. 2021;30(6):685-93. doi:10.1002/pds.5222
29. Vanderpuye J, Gordon M. AI in RWE: Key drivers for accelerating clinical development and patient access. 2024 [cited 2025 Jun 30]. Available from: https://www.ispor.org/conferences-education/conferences/past-conferences/ispor-2024/program/program/session/intl2024-3892/18096
30. Tsai ML, Chen KF, Chen PC. Harnessing electronic health records and artificial intelligence for enhanced cardiovascular risk prediction: A comprehensive review. J Am Heart Assoc. 2025;14(6):e036946. doi:10.1161/JAHA.124.036946
31. Zhao Y, Weroha SJ, Goode EL. Generating real-world evidence from unstructured clinical notes to examine clinical utility of genetic tests: use case in BRCAness. BMC Med Inform Decis Mak. 2021;21(1):3. doi:10.1186/s12911-020-01364-y
32. Kitamura FC. ChatGPT Is shaping the future of medical writing but still requires human judgment. Radiology. 2023;307(2):e230171. doi:10.1148/radiol.230171

Author information

Pattabhi Machiraju, PhD, has been a Senior Director at RWD & Biostats Department of Indegene Ltd since 2011. He leads all AI/ML-driven RWE projects globally and brings over two decades of leadership in biostatistics, real-world data analytics, innovative study designs, regulatory submissions, and cross-functional collaboration with global stakeholders and research partners.

Prarthana Reddy, BTech, MBA, has been a Data Scientist at the RWD & Biostats Department of Indegene Ltd since 2020. With an engineering background, she supports AI/ML-driven RWE projects globally, contributing to algorithm development, data-driven insights, and innovative analytical solutions across various therapeutic areas and RWE initiatives.

Gomathi Priya Jeyapal, PhD, has been a subject matter expert at the RWD & Biostats Department of Indegene Ltd since 2019. She supports AI/ML-driven RWE projects for global stakeholders and brings deep expertise in multiple therapeutic areas, real-world data analysis, evidence generation, protocol development, and cross-functional collaboration.

0000-0003-0529-2121

Vrushabh Satav, BHMS, MBA, has been a domain expert at the RWD & Biostats Department of Indegene Ltd since 2024. He supports AI/ML-driven RWE projects for global stakeholders, bringing strong domain knowledge and expertise in clinical data interpretation across multiple therapeutic areas

Real-world evidence: What does the medical writer need to know?

Harriet S. Crofts, Sarah J. L. Graham
PharmaGenesis London, London, UK

doi: 10.56012/wqvt4437

Correspondence to:
Harriet Crofts
harriet.crofts@pharmagenesis.com

Abstract

Evidence derived from real-world data is invaluable for expanding knowledge about medicines. As medical writers, we need to understand how to think about, handle, and communicate these data to ensure that they are credible and have a meaningful impact. This article shares what we have learned and what we wish we had known when we began working with real-world data.

Interest is growing in what real-world evidence (RWE) adds to medical research. This is not surprising if we consider why medicines are developed and why, as medical writers, we create content to support those medicines. The answer is, of course, to improve the lives of patients receiving care in routine clinical practice whose needs are not currently being met. Although the focus during clinical trials is quite rightly on maximising our understanding of what a medicine is doing by removing as many sources of variance as possible, medicines are ultimately destined for use in highly variable, often unpredictable real-world settings. This places the real world at the heart of the medicine development process. In fact, it is widely recognised today that real-world data (RWD), and the evidence that these data help to generate, complement clinical trials by providing important insights to multiple stakeholders involved across all stages of the medicine lifecycle.[1–3] New guidance is already shaping how RWE studies for use by regulators and payers are conducted to ensure harmonisation of standards, data transparency and reproducibility, as well as to support robustness of study design and the evaluation of bias.[4–7] Integration of RWE within the drug development and approval processes will only increase in the future. Thus, as medical writers, it is increasingly important that, as well as being aware of evolving guidance, we understand how to work with RWD and the unique insights that it can provide. This article outlines how and why RWD differ from data collected within randomised clinical trials and explores the factors that medical writers need to think about when writing about real-world studies.

Medicines are ultimately destined for use in highly variable, often unpredictable real-world settings.

Why do real-world and clinical trial settings differ?

Medicines undergo rigorous testing in tightly controlled clinical trial conditions to ensure their safety and efficacy before they can be used in routine clinical practice. This research is planned in meticulous detail; from the careful selection of individual patients with similar characteristics who fulfil strict inclusion criteria to the use of specific clinical assessments conducted at regular timepoints, everything about the setting is pre-specified. Treatment decisions can also be precisely controlled with randomisation of patients to different treatment groups. While real-world studies can also be planned in detail, they are observational in nature, and neither their environment nor the characteristics of the patients who seek care can be strictly controlled (Table 1).

Treatments and clinical evaluations in the real world are at the discretion of the healthcare professionals who care for the patient. Their decisions are based on clinical experience, local considerations, such as treatment availability and reimbursement, and patient-specific aspects, such as individual goals, lifestyle, and preferences. These factors also influence how often patients see their doctor, undergo tests, and receive prescriptions, not to mention how often they fill their prescriptions and take their medicines. Therefore, while clinical trials typically offer gold standard treatments and regular, detailed assessments for every participating patient, it is unrealistic to expect comparable care in the real world. Furthermore, the documentation of patient care varies between settings, with many real-world sources being unstructured and more variable than the standardised and structured outputs of clinical trials. It is important that medical writers understand these differences between real-world studies and clinical trials and appreciate the strengths and limitations of RWD and the questions it can answer.

What types of research questions can real-world studies address?

Real-world studies, like other forms of medical research, are intended to add value by addressing relevant, unanswered questions. As a medical writer communicating RWE, it is useful to start by considering the research question and the knowledge gap that the study aims to fill, because these will dictate how we introduce it and the context that needs to be provided.

Improving understanding of diseases and patient populations

Real-world studies are often used to improve the understanding of a disease, patient population, and standard of care. For example, the aim of a study may be to quantify how many people are affected by a condition and how this is expected to change over time; to help understand the different stages of a disease and the patient journey; or to characterise the unmet medical needs of patients receiving current treatments. These types of studies can be used to explore the characteristics of any condition and are particularly valuable for expanding our comprehension of rare diseases.[8] Data addressing questions of this kind can also be extracted for use in other studies; a good example is the use of natural history data as an external control group within a single-arm clinical trial.[2]

Table 1. How do real-world studies differ from randomised clinical trials?

	Randomised clinical trial (RCT)	Real-world study
Classification	Interventional	Non-interventional (observational)
Patients	Homogeneous group; often healthier than the average patient in the real world	Heterogeneous group; may have multiple comorbidities and variable disease presentations
Population size	Small relative to the affected population; prespecified based on statistical power for primary outcome	Can be much larger than in RCTs; often unspecified and based on data availability
Doctor/care team	Experienced in the condition being studied; supported by detailed protocols and study team	No guaranteed disease-related experience
Treatments	Pre-specified in protocol; often testing an investigational product; restricted use of concomitant therapies	Local standard of care; dependent on availability and accessibility; wide range of concomitant therapies
Comparisons	Placebo or other medication; designed to detect differences between treatment groups	Often designed to find associations rather than conclude causality
Treatment assignment	Randomised	At the discretion of the treating physician
Treatment blinding	Double/single-blinded or open label	No blinding to treatment
Data collection	Prospective; structured, at consistent time intervals	Retrospective or prospective; often unstructured, at various time intervals
Assessments	Pre-specified in protocol to collect appropriate data for answering study question	Part of routine medical care at variable time points; often not performed with research in mind
Key outcomes	Efficacy and safety	Effectiveness (rather than efficacy) and safety; natural history; disease burden/unmet needs; treatment patterns; costs
Duration	Short; up to several years depending on the research question; limited by feasibility and cost	Can be much longer than RCTs; may cover decades in a retrospective analysis or prospective registry
Study documentation	Protocol, statistical analysis plan, final study report	Level and detail of documentation varies

Exploring treatment use, safety, and effectiveness

Real-world studies can answer questions about the uptake and impact of treatments in larger and more diverse patient populations and over longer time periods than is possible within clinical trials. For instance, a study on treatment patterns may seek to understand who is receiving long-term treatment with a particular medication and whether they are taking that medication as regularly as expected, as well as what other treatments are being prescribed alongside it. Following authorisation of a medicine, real-world studies frequently address questions about safety.[2] These types of investigations allow researchers to detect rare adverse events, monitor risk-management measures, and identify safety signals that warrant further study. Real-world studies can also provide evidence in support of medicine effectiveness by confirming or extending findings of efficacy from clinical trials.

Quantifying disease burden and healthcare impact

The burden of ill health and associated use of healthcare resources are often central themes within real-world studies. Research may measure the impact of a condition and its management by collecting data on direct costs (e.g., for medicines, procedures, and hospital visits) and indirect costs

Box 1. Checklist of questions to ask when writing about a real-world study: aim and audience

Question	Considerations
What is the aim of the study?	Any communication needs a compelling story. • Think about why the study has been done and the real-world question it is trying to answer. • Consider what background information the reader needs to know for them to appreciate the relevance and importance of the study.
Who is the intended audience?	Different audiences will have different areas of expertise and different priorities. • Avoid jargon and explain concepts in simple, unambiguous terms. • Ensure you have the appropriate template when developing documents such as trial protocols and clinical study reports.

(e.g., the impact on ability to work and need for caregiver support), as well as determining any cost savings associated with treatments. Studies may also demonstrate the burden of disease in terms of its impact on the patient's health-related quality of life. Such RWE is important for activities that aim to demonstrate the value of medicines by showing their clinical- and cost-effectiveness or affordability (cost-modelling).[3] For example, RWD may be used in a health economic analysis to help quantify the burden of disease in terms of the number of years of life lost or lived with disability, supporting value comparisons between an existing and future treatment.

Communicating real-world evidence effectively to different audiences

Given the broad range of questions that can be addressed by RWD, a variety of stakeholders are interested in the answers, and therefore it is helpful to consider the needs and expectations of your audience. Medical writers develop many different types of content, including study reports, integrated evidence plans, reimbursement submissions, regulatory documents internal training materials, publications, information for patients, and medical communication materials. The format and audience will influence how RWE is presented. We should also consider how content will be used in the future by different stakeholders. Different audiences will have different areas of expertise – for example, your audience may not be medically trained – therefore, communications need to be tailored appropriately, avoiding jargon and explaining concepts in simple, unambiguous terms. While the language chosen may differ depending on the expertise of the audience, our role as writers is to tell a compelling story that resonates with the reader, whether they are pharmaceutical industry professionals, regulators, payers, healthcare professionals, or patients. Overall, it is important that we think carefully about how to build a narrative that shows our audience(s) how the answer to a real-world question is relevant to them and why the evidence matters. (Box 1)

Understanding data sources and collection methods

To write about a study accurately and highlight its strengths and limitations, we need to understand how it was conducted and why certain methods were chosen over others. Transparent explanations of data source selection and study design are essential when reporting the results of a real-world study.[7,9] Before starting to write a report or publication, it is important to ensure that the source materials that have been provided contain all the necessary study details. If any essential study information is missing or unclear, this should be communicated to the data owner as early as possible, so that the pertinent details can be clarified. Reporting guidelines for non-interventional studies, such as STROBE,[10] which has separate checklists for cross-sectional, case-control, and cohort studies, are useful tools for checking which information needs to be included in a manuscript, and many journals now require the relevant checklist to be completed alongside submissions. Other key guidelines and templates include HARPER which supports the transparent reporting and reproducibility of RWE study protocols.[4] Make sure to read any guidance that applies to your content before starting work.

RWD can be obtained from many different sources, some of the most common being electronic health records, pharmacy and healthcare claims, and product or disease registries (Figure 1), and different types of data come with different strengths and challenges.[11] Databases of medical and pharmacy claims, for example, offer structured data associated with requests for reimbursement for medications or procedures related to a specific diagnosis. In contrast, electronic health records are unstructured but provide much more detail about the health of each patient and the medical care that they received. Patient-reported data, such as responses to surveys or interviews, are highly variable yet give a detailed picture of the true impact of a disease or treatment from the patients' perspective. In order to include comprehensive information in a study, data on individual patients are often combined from different sources, which may be formatted differently and require harmonisation before use. With the increasing availability of large databases of patient information, techniques for converting RWD to RWE are becoming more advanced and incorporating the use of big data, artificial intelligence, and machine learning methods. It is important to bear in mind that the sources used in real-world studies often contain patient data that were not collected with research in mind and may not be fully anonymised, so care must be taken to ensure that these data are reported ethically. Manuscripts should include confirmation of informed consent if appropriate, and either details of ethics committee approval or an explanation of why this was not required.

Data collection outside of the well-defined environment of a randomised controlled trial is inherently variable; therefore, the endpoints used in real-world studies may be more complex than

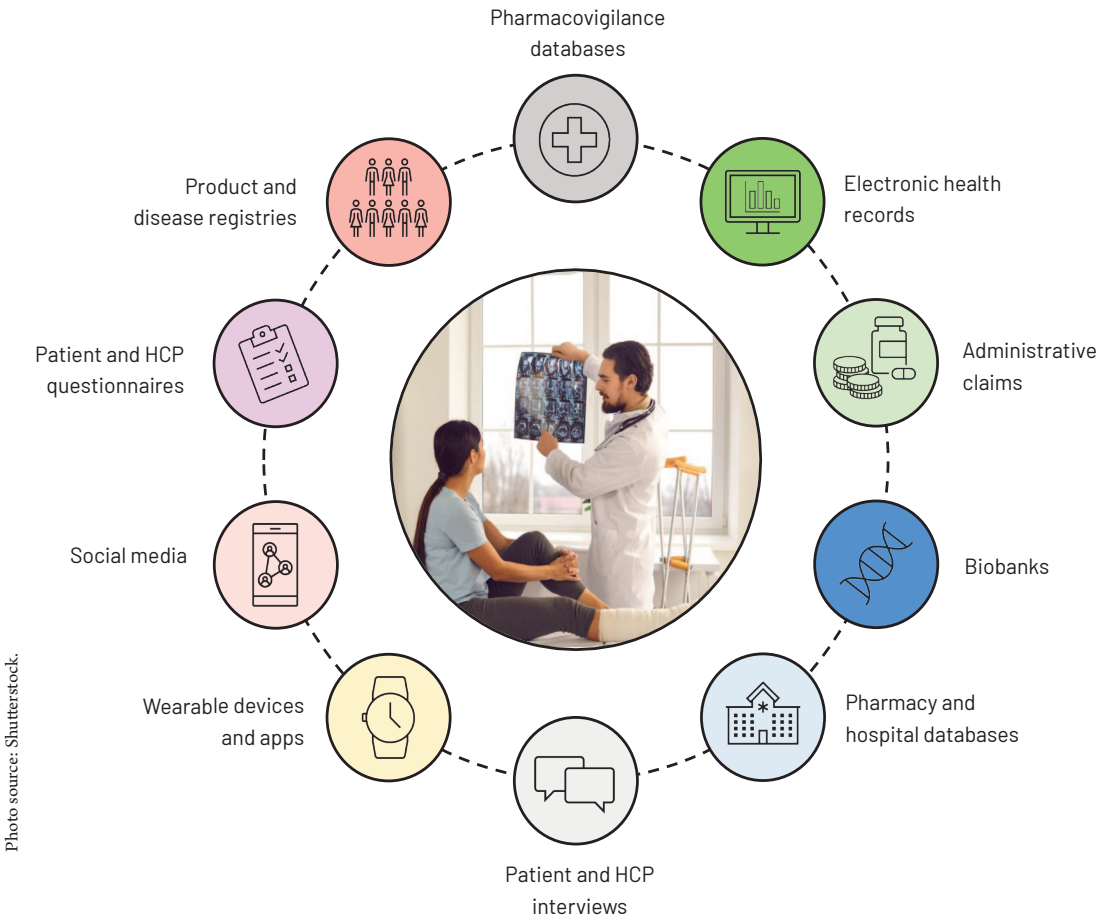


Figure 1. Data sources used in real-world studies

Abbreviations: HCP, healthcare professional

those in clinical trials, particularly if data were collected over long time periods during which diagnostic codes, assessment methods, guidelines, or policies have changed. If data on specific clinical variables are not available, surrogate or composite measures may be used to indirectly determine key outcomes. When reporting outcomes, clarity on the timeframe for follow-up and the patients included in each analysis is essential, as subgroup analyses and missing data are common. As medical writers, we need to think about the nature of the data being reported, how each variable relates to the question we are answering, and what the reader needs to know to understand the study results. (Box 2)

Addressing biases and limitations in real-world studies

Real-world studies have greater potential for bias than randomised controlled trials; therefore, clear reporting of statistical methodology is paramount for building trust that study conclusions are robust. Typical sources of bias in real-world studies include selection bias, information bias, and confounding.[12,13] Selection bias occurs when the selection of individuals or data for a study is not random, and the sample population may therefore not be representative of the wider patient population. This includes self-selection bias, which is relevant when participants choose to be in the sample population, for example, by volunteering to answer an online questionnaire. Information bias arises when key study variables are inaccurately measured or recorded. This includes recall bias, which is a common limitation of studies based on interviews or surveys. Confounding occurs when an uncontrolled variable influences both the independent variable (exposure) and the dependent variable (outcome), so that the results obtained do not accurately reflect the actual relationship between the independent and dependent variables.

Details of the strategies used for minimising bias and handling missing data should be described in the study methods. Common strategies for reducing bias in real-world studies include:[14]

- **Restriction** (strict inclusion and exclusion criteria to create a more homogeneous patient population)
- **Stratification** (dividing the study population into subgroups based on potential confounding variables)
- **Regression analysis** (statistical adjustments using multivariable regression models that take confounding factors into consideration)
- **Propensity score matching** (effectively mimicking randomisation by creating treatment and control groups that are balanced in terms of specific baseline variables)

Sensitivity analyses, which test the potential influence of unmeasured confounders, are also

Box 2. Checklist of questions to ask when writing about a real-world study: source information

Question	Considerations
Do you have all the study details and data you need?	Be prepared/have all the details to hand. • Check guidelines such as STROBE and HARPER for the information that should be reported for real-world studies. The field is evolving rapidly so keep an eye out for new guidance and templates that support data transparency. • Ask the data owner for any missing information as soon as possible.
What data source(s) and study endpoints were used?	RWD sources are numerous and can be very different. • Make sure that you understand the data sources and how they are formatted in enough detail to explain them. • Be clear on how endpoints relate to the study question. • Include sufficient context in the methods and results to allow the reader to assess the data and understand its meaning.
Do any details need to be removed to maintain patient anonymity?	Remember that the data reflect real people. • Keep a look out for details that could compromise anonymity and make sure that they are masked or removed. • Real-world sources such as interviews and free text in questionnaires may contain information that needs to be reported sensitively.
Were patient consent and ethics approval obtained?	Ethics processes for real-world studies may be less straightforward than for clinical trials. • Include some form of statement about ethical review; check with the data owner if the requirements for the study are not clear. • These factors may not be applicable if fully anonymised data were used, in which case, this exemption should be explained clearly.

used to demonstrate the robustness of the results.

As a medical writer developing content based on RWD, it is important to consider the factors that will build confidence in the RWE being presented and to be transparent about potential study limitations. Non-interventional studies are often designed to find associations rather than to conclude causality; therefore, caution with using causal language is needed, particularly when findings are based solely on descriptive statistics. When writing the discussion section of a manuscript, being clear about the generalisability of the results (e.g., that the study only looked at patients of a certain age or from a specific ethnic background) does not diminish the validity of the study but provides essential context for the reader to understand what the results mean. Discussion of limitations is always important, and the inherent limitations of real-world studies should be acknowledged. Any specific limitations identified in study design, data integrity, or interpretation of results should be discussed with the authors to agree on how they should be addressed and whether future studies are warranted. RWD are, by nature, variable and complex; however, real-world studies provide insights that cannot be obtained in clinical trials, and any limitations should be considered in the context of the study's strengths. (Box 3)

Conclusions

The role of medical writers is to develop clear, accurate, and transparent communications, with the ultimate goal of helping to bring evidence-based medicines to patients. When we talk about RWE, it is often caveated with a list of issues that must be addressed, such as data quality, bias, and a general lack of methodological rigor, all of which make it sound challenging. However, RWE is worth the effort – it helps us to fill gaps in clinical trial evidence, offers improved patient-centred insights, lets us look at cost-effectiveness in different geographies, and is receiving growing interest in regulatory and policy circles where it may ultimately help to speed up decision making. By reporting real-world studies effectively and transparently, medical writers can support the pharmaceutical industry in building trust in the diverse and valuable insights gained from RWD.

Acknowledgements

The authors would like to thank Claire Price for reading the article and providing valuable feedback.

Disclaimers

The opinions expressed in this article are the authors' own and not necessarily shared by their employer or EMWA.

Disclosures and conflicts of interest

Harriet S. Crofts is an employee and shareholder of Oxford PharmaGenesis Ltd. The authors declare no conflicts of interest.

Box 3. Checklist of questions to ask when writing about a real-world study: data interpretation

Question	Considerations
What steps were taken to minimise bias?	Real-world studies have more potential for bias than randomised controlled trials. ● Check the protocol and statistical analysis plan (if available) and ask the data owner if unsure. ● Include sensitivity analyses if these were conducted, and results of any other analyses that support data robustness.
Have the outcomes been described appropriately?	Choice of language is important. ● Remember that real-world studies tell us about effectiveness, not efficacy. ● Avoid language around causality/associations if findings are based solely on descriptive statistics.
What are the study strengths?	Non-interventional studies examine what happens in real life in a way that clinical trials cannot. ● Make sure to highlight the strengths of the study, not just the limitations. ● Ask the authors if you are not sure what these are.
What are the study limitations?	Study limitations should be discussed transparently. ● Consider potential sources of bias and the inherent limitations of RWD. ● Discuss potential limitations with the authors to agree on how they should be addressed.
How generalisable are the findings?	The reader needs to understand what the findings mean in a broader context. ● A statement on the generalisability of the study is always important to include. ● Think about the demographics and clinical characteristics of the patients in the study – how representative are they of the global patient population?

References

1. Dagenais S, Russo L, Madsen A, et al. Use of real-world evidence to drive drug development strategy and inform clinical trial design. Clin Pharmacol Ther. 2022;111(1):77–89. doi:10.1002/cpt.2480
2. Franklin JM, Glynn RJ, Martin D, et al. Evaluating the use of nonrandomized real-world data analyses for regulatory decision making. Clin Pharmacol Ther. 2019;105(4):867–77. doi:10.1002/cpt.1351
3. Zisis K, Pavi E, Geitona M, et al. Real-world data: a comprehensive literature review on the barriers, challenges, and opportunities associated with their inclusion in the health technology assessment process. J Pharm Pharm Sci. 2024;27:12302. doi:10.3389/jpps.2024.12302
4. Wang SV, Pottegård A. Building transparency and reproducibility into the practice of pharmacoepidemiology and outcomes research. Am J Epidemiol. 2024;193:1625–31. doi: 10.1093/aje/kwae087
5. Bykov K, Jaksa A, Lund JL, et al. APPRAISE: A tool for appraising potential for bias in real-world evidence studies on medication effectiveness or safety. Value in Health. 2025; in press; doi:10.1016/j.jval.2025.07.024
6. Gatto NM, Vititoe SE, Rubinstein E, et al. A structured process to identify fit-for-purpose study design and data to generate valid and transparent real-world evidence for regulatory uses. Clin Pharmacol Ther. 2023;113:1235–39. doi:10.1002/cpt.2883
7. Barrette E, Crown WH, Hanisch M, et al. Research method, conduct, and reporting considerations for improving the quality of non-hypothesis-evaluating treatment effectiveness analyses using real-world data: An ISPOR special interest group report. Value Health. 2025;28:P979–87. Available from: https://www.valueinhealthjournal.com/article/S1098-3015(25)01967-9/fulltext doi:10.1016/j.jval.2025.02.017
8. Liu J, Barrett JS, Leonardi ET, et al. Natural history and real-world data in rare diseases: applications, limitations, and future perspectives. J Clin Pharmacol. 2022;62(S2):S38–55. doi:10.1002/jcph.2134
9. White R. Building trust in real world evidence (RWE): moving transparency in RWE towards the randomized controlled trial standard. Curr Med Res Opin. 2023;39(12):1737–41. doi:10.1080/03007995.2023.2263353

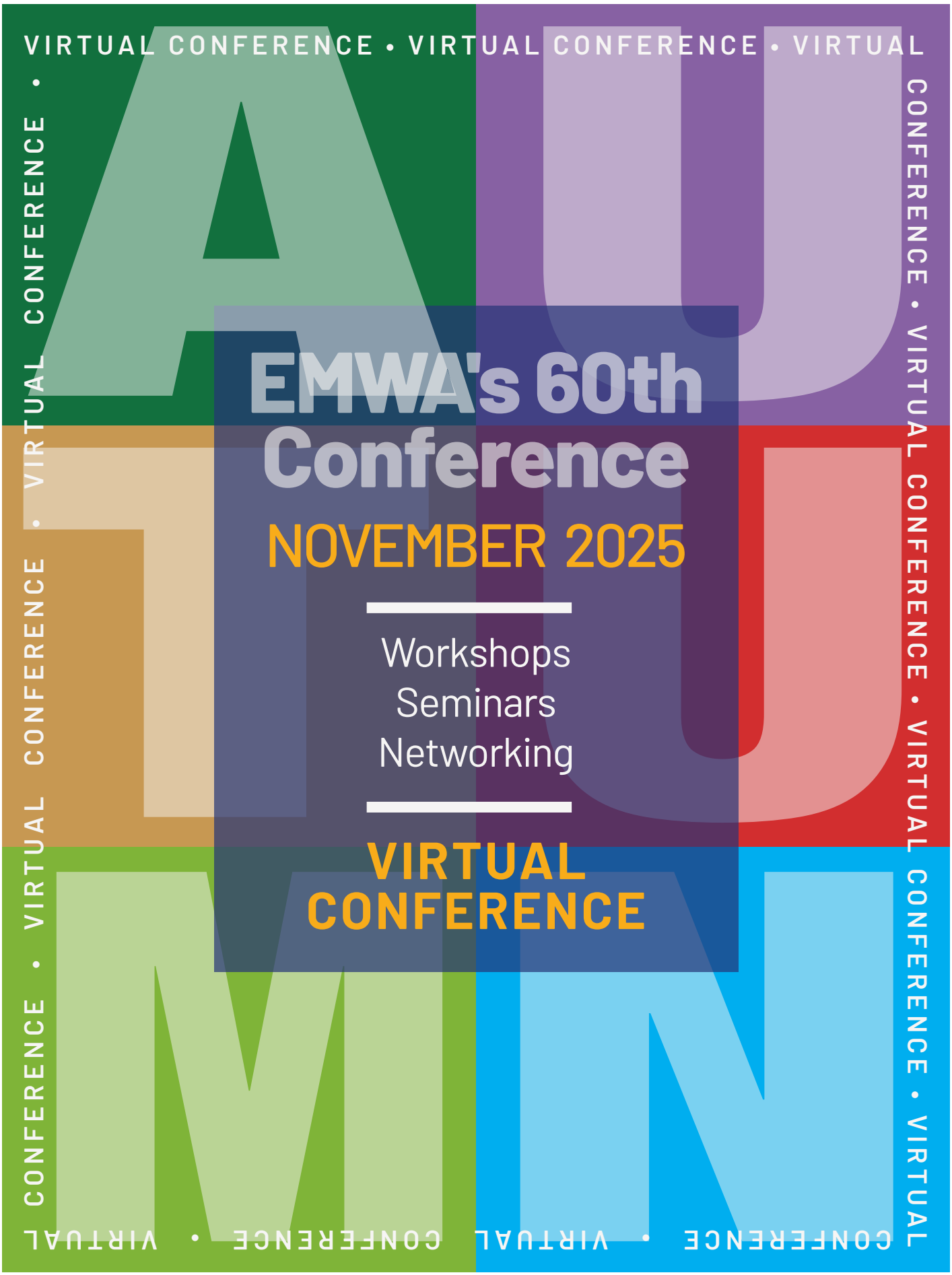
Image: Freepik

10. von Elm E, Altman DG, Egger M, et al. Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. BMJ. 2007;335(7624):806–8. doi:10.1136/bmj.39335.541782.AD
11. Liu F, Panagiotakos D. Real-world data: a brief review of the methods, applications, challenges and opportunities. BMC Med Res Methodol. 2022;22(1):287. doi:10.1186/s12874-022-01768-6
12. EMA. Reflection paper on use of real-world data in non-interventional studies to generate real-world evidence - scientific guideline. June 2025 [cited 2025 August 11]. Available from: https://www.ema.europa.eu/en/reflection-paper-use-real-world-data-non-interventional-studies-generate-real-world-evidence-scientific-guideline.
13. Wang SV, Schneeweiss S. Assessing and interpreting real-world evidence studies: introductory points for new reviewers. Clin Pharmacol Ther. 2022;111(1):145–9. doi:10.1002/cpt.2398
14. Markham JL, Richardson T, Stephens JR, et al. Essential concepts for reducing bias in observational studies. Hosp Pediatr. 2023;13(8):e234–9. doi:10.1542/hpeds.2023-007116

Author information

Harriet S. Crofts, PhD, is a Communications Director at Oxford PharmaGenesis with over 20 years of experience in medical communications. She specialises in rare diseases and has considerable experience working with RWD sources such as registries, questionnaires, and interviews. She has a PhD in Psychopharmacology from the University of Bristol.

Sarah J. L. Graham, PhD, is a Communications Consultant at Oxford PharmaGenesis. She obtained her PhD in Developmental Biology from the University of Cambridge, UK, in 2014 and has been working as a medical writer since 2016, specialising in rare disease publications and medical communications.


VIRTUAL CONFERENCE • VIRTUAL CONFERENCE • VIRTUAL
EMWA's 60th Conference
NOVEMBER 2025
Workshops
Seminars
Networking
VIRTUAL CONFERENCE

Artificial intelligence and machine learning in real-world evidence: Transforming data into actionable insights

Pattabhi Machiraju, Hemalatha Jayapalan, Giles Devasahayam, Manjari Deshmukh
Indegene, Bengaluru, India

doi: 10.56012/rimq2971

Correspondence to:
Giles Devasahayam
giles.devasahayam@indegene.com

Abstract

In recent years, the healthcare industry has experienced an exponential growth in the volume of real-world data (RWD) due to advancements in digital health, electronic health records (EHR), wearables, and other data-generating technologies. The integration of artificial intelligence (AI) and machine learning (ML) into real-world evidence (RWE) generation has the potential to revolutionise how clinical and healthcare decisions are made. AI and ML can effectively analyse large and complex datasets, identifying patterns and insights that were previously hidden or too difficult to detect using traditional analytical methods. For medical writers involved in regulatory submissions, clinical research documentation, and healthcare communications, understanding the application of AI and ML in RWE generation is essential. This publication explores the impact of AI/ML on RWE, its regulatory considerations, and best practices for integrating AI-driven insights into medical writing.

As the healthcare industry shifts towards data-driven decision-making, real-world data (RWD) has become a crucial resource for clinical research, regulatory evaluations, and health technology assessments. The global real-world evidence (RWE) solutions market is projected to reach $4.5 billion by 2029, growing at a compound annual growth rate of 16.9% from 2024 to 2029.[1] This growth reflects the increasing reliance on RWD to enhance patient care, refine treatment strategies, and improve healthcare outcomes. However, ensuring the quality of RWD sources and assessing the feasibility of cross-regional data integration remain key challenges.

RWE has emerged as a transformative force in healthcare, reshaping how medical treatments, interventions, and policies are evaluated.[2] Consequently, the demand for more comprehensive, patient-centred evidence has fuelled the rise of RWE, which leverages data from diverse sources such as electronic health records (EHRs), insurance claims, patient registries, wearable health devices, social media and patient-reported outcome, genomic and biomarker data.[3]

The advancement of RWE has been propelled by innovations in data collection, regulatory endorsement, and analytical methodologies. With the expansion of digital health technologies, vast amounts of patient data have become available, enabling large-scale and more rigorous RWE studies. Regulatory agencies, including the U.S. FDA and the EMA, now acknowledge the value of RWE in complementing traditional evidence for drug approvals, safety monitoring, and health policy decision-making.[4]

Unlocking real-world data: AI and ML's role in actionable insights

Unlocking the potential of RWD for actionable insights is a transformative opportunity. AI and ML are key to overcoming its challenges. Artificial intelligence (AI) and machine learning (ML) enhance data quality by automating error detection, managing missing values, and standardising formats. They also handle unstructured data like text, images, and audio through techniques such as natural language processing and computer vision, enabling analysis of medical records, social media, or sensor data. For data integration, AI resolves entity mismatches and maps diverse datasets to common frameworks, while federated learning allows decentralised training without sharing raw data, ensuring privacy.[5]

AI and ML enable predictive and prescriptive analytics, forecasting trends like disease outbreaks or customer behaviour and recommending optimal actions, such as personalised treatments or dynamic pricing.[6] They also provide real-time insights by processing streaming data from Internet of Things devices or social media, with edge computing enabling instant analysis on devices. To address bias and fairness, AI detects biases in datasets and designs fairness-aware algorithms, while synthetic data generation helps reduce biases and protect privacy.[7]

AI also supports personalisation, powering recommendation systems in retail or healthcare, and enhances scalability through automated machine learning (AutoML) and distributed computing. Privacy-preserving techniques like federated learning and differential privacy ensure secure data usage. Finally, explainable AI (XAI) tools make models transparent, building stakeholder trust and enabling actionable insights. By leveraging AI and ML, organisations can transform RWD into meaningful, innovative solutions across industries.

Applications of AI/ML in RWE

Predictive modelling for patient outcomes and disease progression

AI/ML can be applied to large-scale RWD datasets to forecast patient outcomes and understand disease progression. This predictive capability is crucial for personalising treatment strategies and identifying at-risk patients who may benefit from early intervention. Predictive models can estimate the likelihood of disease complications, hospital readmissions, or treatment responses, enabling healthcare providers to proactively tailor interventions.[9] This not only improves patient outcomes but also reduces unnecessary healthcare costs. Additionally, these models help in predicting disease progression, allowing clinicians to identify high-risk patients who may need more aggressive interventions.

For example, in patients with diabetes, AI/ML algorithms can detect patterns indicative of impending complications such as diabetic retinopathy, nephropathy, or cardiovascular disease by analysing trends in haemoglobin A1c levels, blood pressure readings, and renal function markers. Similarly, in patients with heart disease, predictive models can assess risk factors such as cholesterol levels, previous cardiac events, medication adherence, and electrocardiogram findings to estimate the likelihood of heart failure exacerbations or myocardial infarctions.[10]

Identifying treatment patterns and healthcare resource utilisation trends

AI/ML are powerful tools for uncovering trends in treatment patterns and healthcare resource utilisation. By analysing large-scale datasets, these technologies can track how treatment approaches evolve, identify the most effective real-world therapies, and reveal care patterns across diverse patient groups. For instance, ML can examine EHRs to determine the most frequently prescribed treatments for specific conditions, enabling clinicians and policymakers to evaluate the effectiveness of current practices. These insights can also support cost-effectiveness analyses, guiding healthcare systems in optimising resource allocation.[11]

Analysing RWD on hospital admissions and discharge rates can offer valuable insights into the effectiveness of post-surgery care plans. By monitoring readmission rates and length of stay, hospitals can evaluate whether patients receive adequate post-discharge care and pinpoint risk factors leading to complications. Identifying common causes of readmissions, such as infections or medication non-adherence, allows for targeted interventions to improve care coordination and prevent unnecessary hospitalisations. Predictive analytics can further enhance resource utilization by optimising staffing, bed management, and follow-up care strategies. Ultimately, leveraging RWD refines post-surgical care, reduces readmission rates, and improves both patient outcomes and healthcare system efficiency.[12]

Generative AI (GenAI) in generating RWE to inform and optimise clinical trial design using RWD

GenAI helps enhance the design of a clinical trial (e.g., inclusion/exclusion criteria, endpoints, comparator arms) by leveraging RWD such as EHR, claims data, and patient registries, transformed into actionable real-world evidence (RWE) leading to faster patient recruitment, cost reduction, dynamic protocol adjustments and higher external validity (Table 1). GenAI models process large-scale, unstructured RWD (e.g., clinician notes, lab results) aiding data ingesting and harmonization.[13] On the basis of target disease profiles, GenAI simulates patient cohorts from RWD to reflect "real-world" populations, identifies characteristics such as comorbidities, treatment patterns, and geographic distribution, which helps simulate patient cohorts.[14] In comparative effectiveness modelling, GenAI generates synthetic control arms using matched historical patient data, reducing or replacing the need for placebo groups and assesses comparative effectiveness of drugs, aiding in endpoint selection.[15] In trial feasibility and site

Table 1. AI/machine learning applications in real world evidence generation across healthcare

Use case	Application	Data source	AI/ML role	Benefits
Advancing cancer research and drug development	Processing genomic data to identify therapy-linked mutations	Genomic profiles from cancer patients	Analyse genetic mutations to develop targeted, personalised treatments	Enhanced treatment precision, improved outcomes
Forecasting hospital readmissions	Predicting which patients are at high risk of readmission	EHRs: comorbidities, past hospital stays, labs, social determinants	Identify high-risk patients by evaluating clinical and non-clinical risk factors	Lower readmission rates, better care quality, reduced healthcare costs
Monitoring drug safety in real-time	Continuous pharmacovigilance post-marketing	EHRs, insurance claims data	Detect adverse events and safety signals in near real-time	Improved patient safety, proactive regulatory decisions
Patient stratification and disease progression	Grouping patients by risk level and predicting disease progression	Clinical records, genetic markers, drug use, adherence, cost data	Stratify patients into risk categories and forecast disease trajectory	Better outcomes, personalised care, efficient resource use

Abbreviations: EHR, electronic health records; ML, machine learning

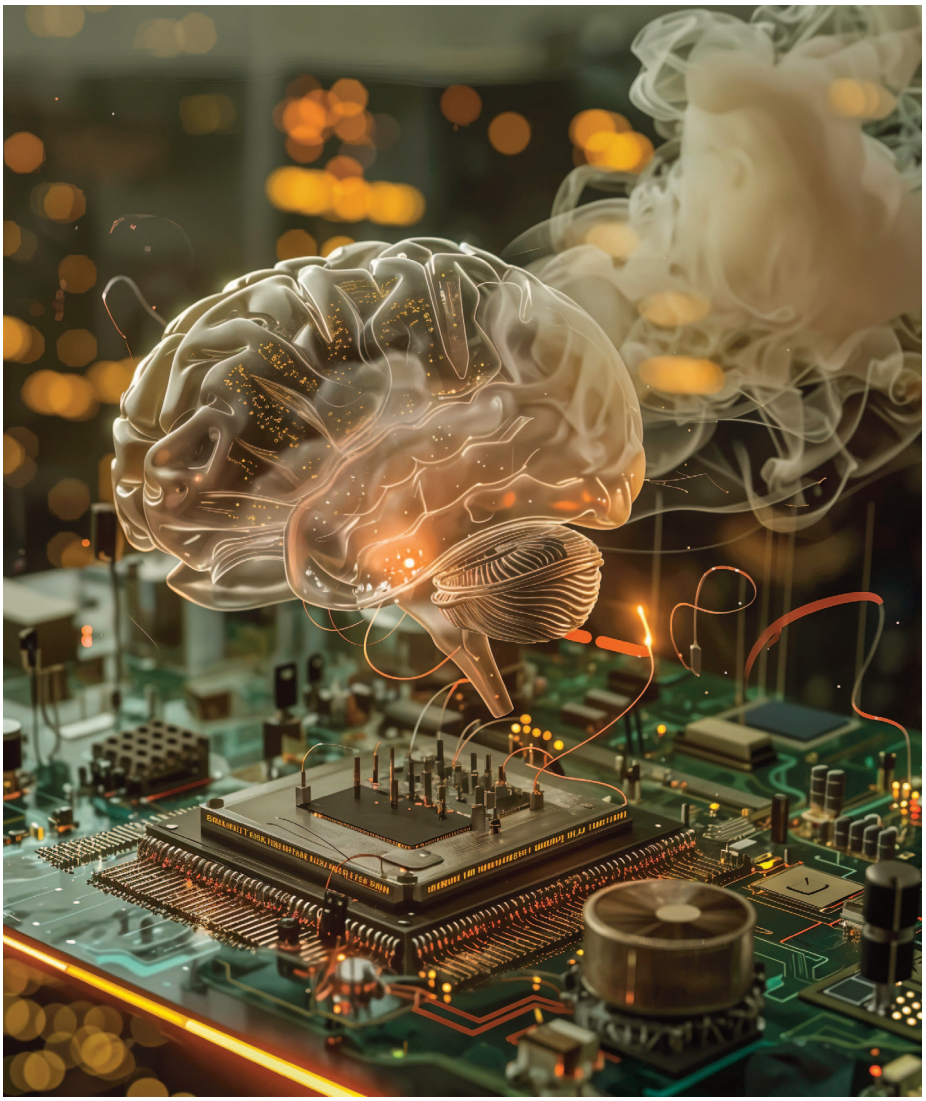


Image: Freepik

selection process GenAI models use RWE to predict patient availability by geography and site, optimising site selection and reducing recruitment delays process.[16,17] GenAI can simulate synthetic patient populations that reflect real-world demographics and disease patterns, in synthetic data generation process which enables simulated control arms for clinical trials and pre-trial feasibility assessments.[18] GenAI has been significantly used in protocol optimisation and feasibility studies to generate and refine trial protocols by predicting dropout rates, identifying high-performing sites, and assessing protocol feasibility based on RWD.[19]

Challenges and ethical considerations

Tackling bias in data and algorithms

A major challenge in applying AI and ML to RWE is ensuring the training data are unbiased. RWD often lack representation of certain populations, which can result in skewed outcomes and disparities in healthcare. When AI models are trained on such imbalanced data, they may reinforce these biases, producing less accurate predictions for underrepresented groups.[20]

Some recommended approaches that could be used to resolve the challenge are:

a. Using diverse and representative datasets for training AI/ML models, ensuring that all demographic groups are adequately represented.
b. Performing regular audits of algorithms to identify and mitigate biases.
c. Utilising fairness-aware ML techniques that are designed to reduce bias during model development.

Promoting transparency and interpretability in AI/ML models enhancing transparency and interpretability in AI/ML models

The lack of transparency in many AI systems – particularly those based on deep learning – poses a significant concern in healthcare, where it's essential to understand how decisions are made. Clinicians, patients, and regulators require insight into the reasoning behind model outputs, but this is often hindered by the so-called "black box" nature of complex models.

A few approaches could foster transparency and interpretability in AI/ML models:

a. **Implementing explainable AI (XAI) approaches to shed light on model decision-making processes.** Techniques like feature importance analysis and decision trees can offer valuable insight into how predictions are generated, helping stakeholders grasp the underlying rationale.
b. **Designing intuitive interfaces that clearly convey model outputs** to healthcare providers and patients, making complex results easier to interpret and use in decision-making.
c. **Offering clear documentation and training resources** to educate users about the assumptions, limitations, and appropriate use of AI/ML models.
d. **Foster collaboration among data scientists, healthcare professionals, and regulators** to ensure models are both technically robust and aligned with clinical needs.

The role of medical writers in interpretation and dissemination of RWE insights derived from AI/ML

Simplifying AI/ML insights for clinical decision-making

Medical writers act as a bridge between the technical world of AI/ML and the practical needs of clinicians. They take complex outputs – such as predictive models, risk scores, or pattern recognition – and translate them into clear, actionable insights. By focusing on how these insights can improve patient outcomes or streamline clinical workflows, medical writers ensure that AI/ML findings are not just understood but also applied effectively in real-world healthcare settings.

Communicating RWE with precision and transparency

When medical writers interpret RWE and generate insights, they must take several precautions to ensure accuracy, relevance, and scientific integrity. RWE can be powerful but also complex and prone to misinterpretation if not handled rigorously. Medical writers play a key role in presenting these findings accurately and clearly. They explain the data sources, methodologies, and limitations in a way that is both scientifically rigorous and easy to understand. By providing context and highlighting the most relevant takeaways, medical writers help clinicians and stakeholders trust and utilise AI/ML-driven RWE in their decision-making.

Delivering actionable insights without compromising integrity

Medical writers ensure that AI/ML-driven insights are presented in a way that is both practical and ethical. They focus on how the findings can be used to improve patient care or inform policy, while also addressing the limitations of the data and models. By avoiding overhyped claims and emphasising the need for responsible interpretation, medical writers help maintain the credibility and scientific integrity of AI/ML applications in healthcare.

Conclusion

The integration of AI/ML into RWE is revolutionising healthcare decision-making by transforming vast amounts of data into actionable insights. The application of AI/ML methods to RWD enables the identification of patterns, prediction of outcomes, and simulation of interventions at scale. However, the complexity of these methodologies poses significant communication challenges. Medical writers are key to ensuring that the resulting insights are both accessible and meaningful to diverse stakeholders. As AI/ML-derived RWE becomes increasingly central to healthcare innovation, the role of the medical writer is evolving from content developer to strategic partner. By ensuring transparency, accuracy, and accessibility, medical writers enable stakeholders to act on complex insights with confidence.

Disclaimers

The views expressed in this article are the personal opinions of the authors and do not necessarily reflect the position of the authors' affiliated organisations.

Disclosures and conflicts of interest

The authors declare no conflict of interest.

References

1. Real world evidence solutions market: growth, size, share, and trends. [AW1] [VW2] Markets and Markets. 2024 [cited 2025 Feb 17]. Available from: https://www.marketsandmarkets.com/Market-Reports/real-world-evidence-solution-market-76173991.html
2. Kaatje B, Chloe WT, Miller E, Izurieta HS, Black S, Andrews N, et al. The role of real-world evidence for regulatory and public health decision-making for Accelerated Vaccine Deployment- a meeting report. Biologicals. 2024;85:101750. doi:101016/j.biologicals.2024.101750
3. Dang A. Real-world evidence: A primer. Pharm Med. 2023;37(1):25–36. doi:10.1007/s40290-022-00456-6
4. Shaikh J. AI Revolution in Evidence Synthesis and Research: NICE Insights. Axtriacom. 2024 Sep 23 [cited 2025 Feb 17]. Available from: https://insights.axtria.com/articles/revolutionizing-evidence-synthesis-with-ai-insights-from-nices-latest-position-statement
5. Chen C, Feng X, Li Y, et al. Integration of large language models and federated learning. Patterns (N Y). 2024;5(12):101098. doi:10.1016/j.patter.2024.101098
6. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. N Engl J Med. 2019;380(14):1347–58. doi:10.1056/NEJMra1814259
7. Arora A, Wagner SK, Carpenter R, et al. The urgent need to accelerate synthetic data privacy frameworks for medical research. Lancet Digit Health. 2025;7(2):e157-e160. doi:10.1016/s2589-7500(24)00196-1
8. Froling E, Rajaeean N, Hinrichsmeyer KS, et al. Artificial intelligence in medical affairs: A new paradigm with novel opportunities. Pharmaceut Med. 2024;38(5):331–342 doi:10.1007/s40290-024-00536-9
9. Kragelund SH, Kjarsgaard M, Jensen-Fangel S, et al. Research electronic data capture (REDCap®) used as an audit tool with a built-in database. J Biomed Inform. 2018;81:112–8. doi:10.1016/j.jbi.2018.04.005
10. Gunasekeran DV, Ting DSW, Tan GSW, et al. Artificial intelligence for diabetic retinopathy screening, prediction and management. Curr Opin Ophthalmol. 2020;31(5):357–65. doi:10.1097/ICU.0000000000000693
11. Thomas R, Chalkidou K. Cost–effectiveness analysis. In: Cylus J, Papanicolas I, Smith PC, editors. Health system efficiency: How to make measurement matter for policy and management. . Copenhagen (Denmark): European Observatory on Health Systems and Policies; 2016. (Health Policy Series, No. 46.) 6.
12. Rajkomar A, Oren E, Chen K, et al. Scalable and accurate deep learning with electronic health records. NJP Digit Med. 2018;1(1):18. doi:10.1038/s41746-018-0029-1
13. IbrahimM, Al Khalil Y, Amirrajab S, et al. Generative AI for synthetic data across multiple medical modalities: A systematic review of recent developments and challenges. Comput Biol Med. 2025;189:109834. doi:10.1016/j.compbiomed.2025.109834
14. Beam AL, Kohane, IS. Big data and machine learning in health care. JAMA. 2018;319(13):1317–18. doi:10.1001/jama.2017.18391
15. Kalidindi VR, Rehana S, Seethamraju SM, et al. Revolutionizing medicine: Unleashing the power of real-world data and AI in advancing clinical trials. Braz J Pharm Sci. 2024:60:e23980. doi:10.1590/s2175-97902024e23980

Image: Freepik

16. Sachdeva S, Kaneria J, Malik R, et al. MSR40 leveraging artificial intelligence (AI) and generative AI (GenAI) for transforming real-world evidence (RWE)across the product value chain and industry functions. Value Health. 2024;27(12):S445. doi:10.1016/j.jval.2024.10.2274
17. Chopra H, Shin DK, Munjal K, et al. Revolutionizing clinical trials: the role of AI in accelerating medical breakthroughs. Int J Surg. 2023;109(12):4211-–20. doi:10.1097/JS9.0000000000000705
18. Yoon J, Jarrett D, Van der Schaar M. Time-series generative adversarial networks. Adv Neural Inf Process Syst. 2019:32.
19. Chatterjee S, Fruhling A, Kotiadis K, et al. Towards new frontiers of healthcare systems research using artificial intelligence and generative AI. Health Syst (Basingstoke). 2024;13(4):263–73. doi:10.1080/20476965.2024.2402128
20. Markus AF, Kors JA, Rijnbeek PR. The role of explainability in creating trustworthy artificial intelligence for health care: A comprehensive survey of the terminology, design choices, and evaluation strategies. J Biomed Inform. 2021;113:103655. doi:10.1016/j.jbi.2020.103655

Author information

Pattabhi Machiraju, PhD, is a Senior Director at RWD & Biostats department, Indegene. He leads all AI/ML-driven RWE projects globally and brings over two decades of leadership in biostatistics, real-world data analytics, innovative study designs, regulatory submissions, and cross-functional collaboration with global stakeholders and research partners.

Manjari Deshmukh, PG in Bioinformatics, serves as a Lead in Analytics at RWD and Biostatistics. With over 11 years of experience in healthcare, pharmaceuticals, and genomics data, she specialises in advanced data analytics. Her expertise encompasses both qualitative and quantitative studies, focusing on real-world data and RWE.

Dr Hemalatha Jayapalan, BDS, PGDCR, is a Senior Manager at RWE & HEOR, Indegene. She is a Subject Matter Expert leading the RWE & HEOR operations. With over 15 years of experience in the Healthcare and Pharmaceutical industry, she possesses extensive expertise in designing, implementing, and executing evidence generation strategies that support market access, reimbursement, and regulatory submissions.

Giles Devasahayam, M. Pharm, PhD, is a Lead in Analytics – Bioinformatics at Indegene, bringing over 15 years of expertise at the intersection of drug discovery, translational bioinformatics, and real-world data. As a subject matter expert, he leverages advanced analytics techniques to generate actionable insights that drive clinical development and commercial strategy.

Don't miss!

The March 2026 edition

Medical Writing Careers

In this issue, we are exploring careers within medical writing, including getting that first role, internship programmes, different types of medical writing, and how a medical writing career may not go in the direction you expected. We will look at how to become more specialised, how to move into management, and potential opportunities after you stop writing.

Guest Editors: Andrew Balkin and Jules Kovacevic
The deadline for feature articles is December 1, 2025.

The paradigm of real-word evidence in lifestyle medicine: Insights, considerations, and opportunities for medical writers

Maria Carolina Rojido,[1]
Mariana Rickmann,[2] **Laura A. Kehoe**[3]

[1] Freelance Medical Communications and Scientific Writing Consultant, Colomiers, France
[2] Freelance Medical, and Scientific Communications, Olching, Germany
[3] Medical Communications & Coaching, Hauterive, Switzerland

doi: 10.56012/nyoc9654

Correspondence to:
Maria Carolina Rojido
carolinarojido@gmail.com

Abstract

Non-communicable diseases are rising at an alarming rate across the globe, with many attributed to our sedentary habits, unhealthy diets, chronic stress, poor sleep, and social, and environmental factors. Lifestyle medicine is an evidence-based discipline that has the potential to prevent, treat, and sometimes reverse chronic illnesses by addressing modifiable lifestyle factors through behavioural interventions. This article highlights the importance of real-world data to objectively evaluate outcomes and advance research in lifestyle medicine. We explore the current literature and the characteristics of the body of evidence on lifestyle interventions and provide tips for medical writers when working with this type of data.

Lifestyle medicine and its potential

With the growing prevalence of non-communicable diseases (NCDs), such as obesity, type 2 diabetes, cardiovascular disease, and cancers, researchers and policymakers are recognising the significant role our lifestyle habits and environmental factors play in this global health burden.[1,2] According to the WHO, 80% of deaths related to NCDs are linked to key modifiable lifestyle-related risk factors: tobacco use, physical inactivity, unhealthy diet, and the harmful use of alcohol.[3,4]

Lifestyle medicine (LM) is emerging as a practical and efficacious medical approach to manage/address/contain the NCD epidemic.[5] It is an evidence-based medical discipline that targets daily habits to address the root causes of health conditions, thereby preventing, treating, and sometimes reversing many chronic diseases that affect people worldwide. Some of the official definitions of LM from established national and international organisations are provided in Table 1.

> Lifestyle medicine addresses the risk factors of chronic diseases, rather than simply alleviating symptoms and signs. It emphasises behavioural interventions to achieve sustainable health improvements.

As we can see, LM comprises specific pillars:[6]

- Nutrition
- Physical activity
- Stress management
- Restorative sleep
- Social connection
- Avoidance of risky substances (drugs, tobacco, alcohol)

Two extra pillars have been recognised by the European Lifestyle Medicine Organisation (ELMO):

- Sexual health and fertility
- Environmental exposure

In line with recent and fast-growing initiatives in the patient engagement space, which in general invite patients to be more involved in their health journey, LM actively engages the patient as a partner in care and decision-making. LM coaching consists of a collaborative process based on motivational interviewing to recognise issues and habits and to empower patients to improve their health through behavioural interventions.[7] LM should not be confused with integrative or alternative therapies, or therapies such as acupuncture or nutraceuticals.[8] Equally, pharmacological treatments are not necessarily excluded from LM interventions; in fact, they may be necessary in many cases for different reasons. Thus, lifestyle and conventional medicine complement each other in clinical practice and in research studies.

Unfortunately, many of us lead sedentary lives and eat westernised diets. There is (still) no pill to replace healthy lifestyles. Despite recent obesity drugs making dramatic improvements for some people, they are not an option for everyone and they are not free of side effects.[9]

Randomised clinical trials (RCT), as the gold standard for clinical research, make drug development possible and safe. But our daily habits have long-term and multifactorial effects on our health – influenced also by social, physical, and mental health factors – which require different methods.[10] In addition, as LM is an evidence-based discipline, real-world data (RWD) and real-world evidence (RWE) are essential to drive LM forward.[11,12] RWD provide substantial data to measure outcomes and RWE provides insights into the effectiveness of lifestyle interventions in diverse populations over

Image: Freepik

extended periods. There is a general belief that all evidence in LM stems from RWD; however, clinical trials like the PREDIMED trial can also explore the outcomes of lifestyle interventions.[13,14]

To evaluate how much of LM evidence relies on RWD, we conducted a simple literature search to identify the RWE on lifestyle interventions and the characteristics of the body of evidence in this discipline. We present our findings below. We also provide information on data sources and useful tips that medical writers (MWs) could apply when working in LM research.

Exploring published and current clinical studies in LM

To explore the current state of research and publications on LM and its pillars, we conducted searches in clinicaltrials.gov and PubMed. We acknowledge, as a limitation of this preliminary search, that some of the studies found in clinicaltrials.gov may be published and thus duplicated in the search in PubMed. Although these searches do not constitute a formal review, the aim was to explore the evidence objectively and provide MWs with concrete information to help them understand this field and be better prepared when they encounter these topics in their work. Please see the Appendix for a detailed description of the methods we used.

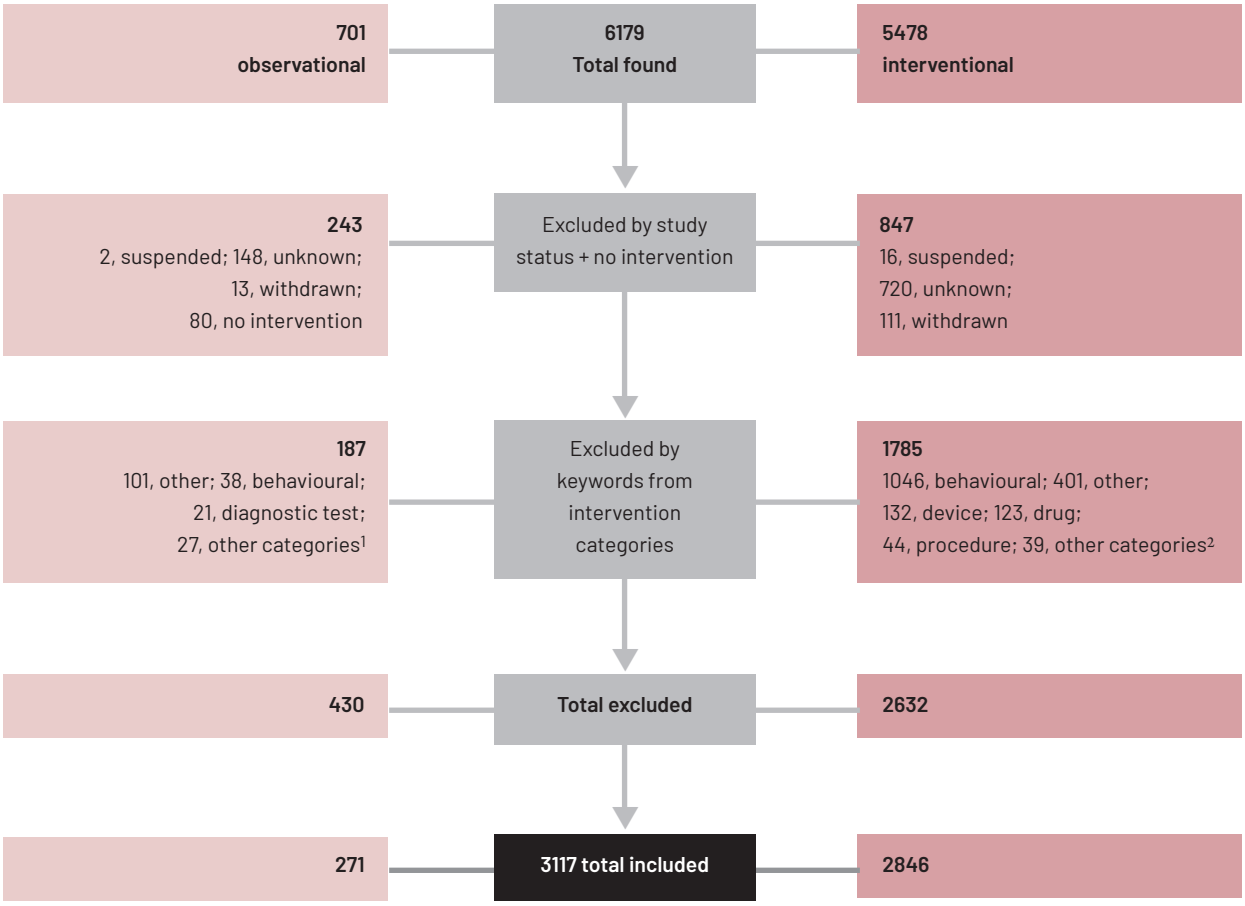
Our clinicaltrials.gov search yielded 6179 studies: 701 observational and 5478 interventional.

Table 1. Official definitions of lifestyle medicine

Belgian and European LM Organisation (BELMO or ELMO)	LM is a branch of medicine that has the goal to maintain optimal health and to **prevent, treat, and reverse chronic illness** across all life stages. The health interventions used in LM include evidence-based behavioural strategies, while considering equity, and sustainability, to enhance self-management skills for optimising nutrition, sleep hygiene, stress management, social connection, sexual health, fertility, and physical activity, and minimising substance use and environmental exposures.
British Society of LM (BSLM)	LM is evidence-based, clinical care that supports behaviour change through person-centred techniques to improve mental wellbeing, social connection, healthy eating, physical activity, sleep, and minimisation of harmful substances and behaviours. It acknowledges the need for action on socioeconomic determinants of health, provides education around the six key pillars as well-proven techniques to sustain lifestyle changes. To be an effective antidote to the chronic disease problem, LM requires a multidisciplinary multi-system approach — which embraces and works alongside other approaches such as self-care, self-management, social prescribing, and group consultations. It requires clinicians, public health professionals, researchers, scientists, and educators working together to affect change. The principles of LM must be applied not only at the clinical practice level, but must also encompass public health policy and prevention. Healthcare professionals, individuals, and governments, and policy makers must play their part.
American College of LM (ACLM)	LM is a medical specialty that uses therapeutic lifestyle interventions as a primary modality to treat chronic conditions including, but not limited to, cardiovascular diseases, type 2 diabetes, and obesity. LM certified clinicians are trained to apply evidence-based, whole-person, prescriptive lifestyle change to treat and, when used intensively, often reverse such conditions. Applying the six pillars of LM – a whole-food, plant-predominant eating pattern, physical activity, restorative sleep, stress management, avoidance of risky substances, and positive social connections – also provides effective prevention for these conditions.

Abbreviation: LM, lifestyle medicine

Figure 1. Flowchart showing the study filtering process after preliminary search in clinicaltrials.gov using "lifestyle" as the intervention



After the exclusion process shown in Figure 1, we selected 3117 studies: 271 observational and 2846 interventional. In both categories, the studies aimed to treat predominantly NCDs (obesity, 660; diabetes, 426; cardiovascular, 323; cancer, 230) but many other conditions as well (neurologic, 183; mental /stress, 100; female, 71; sleep, 62). Among the 2846 interventional studies, 2342 were randomised, and 1308 of these 2342 studies had some type of masking (mostly single, but also double, triple, etc.). Notably, fewer than one in ten of the study postings (268/3117) included study results.

Over a third of the studies took place in the US followed by the sum of countries located in Europe (Table 2). Two-thirds of them consisted of behavioural interventions followed by "other" which were mostly lifestyle-focused, dietary supplements, drugs, and devices. Among the interventional studies, close to half of them had a single lifestyle/pillar-focused intervention and over half of them consisted of ≥2 interventions, where one was always a lifestyle/pillar-focused intervention and the others a different lifestyle/pillar-focused intervention or varying combinations of drugs, combination products, devices, diagnostic tests, procedures, etc.

Close to two-thirds of all the studies were focused on lifestyle; studies focused on the other pillars were less common (nutrition, physical activity, stress, and sleep). But considering that we searched using a unique keyword (lifestyle, under intervention), we obtained fewer studies on individual pillars. We did not include pillar-specific keywords in our search.

It is important to note that this is a preliminary search. Nevertheless, we found that lifestyle-related interventions are being included in research for multiple conditions as lone interventions or as valid comparators. In addition, we obtained more interventional than observational studies; we speculate that this could be because interventional studies should be registered in a clinical trials registry to be considered for publication (per many journal editorial policies) and thus, would bias the results we have obtained.

A potential weakness of our review is that clinicaltrials.gov study postings rarely include results. This is a well-reported problem regardless of the topic under investigation.[15,16] Although it would be logical to hypothesise that the results of the two searches may overlap, it is worth noting that not all published studies are posted in clinical trial registries and that clinicaltrials.gov is not the only registry of this kind. Future reviews should include manual verification of the

Table 2. Characteristics of interventional and observational studies found in clinicatrials.gov preliminary review

	Interventional	Observational
Intervention		
Behavioural	1917	82
Other (mostly lifestyle-focused)	481	75
Dietary supplement	150	4
Drug	140	8
Device	94	8
Procedure	42	6
Other categories	22	9
TOTAL	**2846**	**271**
Pillar		
Lifestyle, coaching, habits, behaviour, counselling, motivational	1756	204
Nutrition, diet	436	40
Exercise, physical activity	487	17
Sleep	67	6
Stress, mindfulness, meditation	100	4

	Interventional	Observational
Country		
United States	1132	56
Canada	123	12
UK	80	14
Spain	136	11
Italy	80	19
Netherlands	66	7
Germany	64	5
Sweden	63	1
Norway	58	4
France	44	17
Start Year		
2020–2025	1187	124
2015–2019	688	72
2010–2014	539	37
2005–2009	299	20
2000–2004	102	9
<1999	21	9

Abbreviations: CT, clinical trial; RCT, randomised clinical trial

Table 3. Types of articles and their publication years found in PubMed that had in their titles keywords representative of LM and four of its pillars

	Lifestyle	Nutrition	Physical activity	Stress	Sleep
CT or RCT	2520	17,423	24,997	11,745	7394
Observational	316	2103	1750	1163	1598
Case report	125	3635	2275	3258	3029
Review	3053	45,627	22,812	31,460	15,031
Systematic reviews and meta-analyses	862	7021	9541	4073	3807
Editorial	586	6359	3256	3232	3116
Letter	487	63,640	4041	3646	3739
Practice guideline and guidelines	27	750	242	82	167
TOTAL	**7949**	**145,808**	**68,672**	**58,577**	**37,714**
Year of publication					
<1999	632	20,076	9548	14,366	5811
2000–2004	1214	25,562	12,697	22,201	8070
2005–2009	2351	36,054	20,537	37,841	12,474
2010–2014	4494	57,434	33,428	60,563	20,394
2015–2019	6469	83,225	47,445	83,548	29,959
2020–2025	11,712	136,164	68,910	129,566	49,958

Abbreviations: CT, clinical trial; LM, Lifestyle Medicine; RCT, randomised clinical trial

overlap between published studies and results posted in clinicaltrials.gov, to better understand its magnitude.

Our PubMed search results showed similar number of studies for lifestyle and four of its pillars (7949): nutrition, physical activity, sleep, and stress (Table 3). Unlike in the clinicaltrials.gov search, the PubMed search enabled us to look individually at lifestyle and each of the selected pillars. We found that the overall pattern was the same: observational studies were only a small percentage compared to clinical trials (7.0% to

21.6%). This is remarkable because the usual paradigm is that RWE predominates in LM.

A large proportion (38.4%) of the other article types was reviews (Table 3), followed by systematic reviews and meta-analyses (10.8%). Notably, practice guidelines and guidelines constituted a very small proportion (0.034%). Yet, considering that we sought only articles with the specific lifestyle-related keyword in the title of the article, the existence of guidelines denotes significant and focused efforts to incorporate LM and its pillars into clinical practice.

Our two searches showed that the number of studies on LM and some of its pillars have increasing trends across each 5-year period studied. For example, in 2015–2019, 760 studies were started, and 6469 articles were published; while in 2020–2025, these increased to 1311 and 11712, respectively. Even if we subtracted those from the first quarter of 2025, the increase is important (Tables 2 and 3).

It should be noted that both searches, despite their preliminary nature, yielded consistent results. The increasing number of studies and publications over time on these topics matches prior findings that highlighted how work in and around LM and its pillars is steadily and rapidly increasing, even compared with the number of studies published in oncology (see Rojido MC, *Medical Writing*, 2019).[8]

What type of LM RWD can we gather and from where?

We can see that there is a growing trend of studies and publications around lifestyle interventions; but, where does this data come from? How do researchers gather RWD to analyse the lifestyle interventions in interventional and real-world settings?

RWD provide a rich source of insights for researchers to analyse health outcomes in non-controlled, everyday settings, and can be transformative in LM. Even more so, we are in an era where digital health is literally at our fingertips, on our smart phones, watches and rings, and these sources are now advancing with the development of sophisticated AI and machine learning programmes. Thus, there are various methods to gather LM-related health data. Here are just a few:

- **Wearable biometric devices:** A vast amount of data is being generated and shared by fitness and health trackers, such as smartwatches, bracelets, or rings. These are often AI-enabled wearable biometric devices and sensors that continuously monitor health metrics like heart rate, blood pressure, and sleep patterns, and can alert the wearer to abnormalities detected. They have a two-way facet: firstly, they assist users in pursuing a healthier lifestyle and in being in control of their health, and secondly, they can provide a constant stream of data for health and safety monitoring, chronic disease management, disease diagnosis, and treatment and rehabilitation.[17] Data is usually saved on the device or the smartphone app and uploaded to servers so users can access their health data across devices. If a user selects third-party services, then the data will be shared (usually anonymously) with health research platforms, the user's healthcare provider, or other providers.
- **Mobile health (mHealth) technologies:** Similar to the wearable devices, users can input their data into mHealth apps on smartphones, web-based technologies, and telecommunications or telemedicine services and log physical activity (steps, exercise minutes, GPS-traced walks or runs, etc.), nutrition (food logs, photo-based meals, calorie counts, etc.), sleep (duration, quality, bedtime, wake time, self-reported restfulness, etc.), mental wellbeing (mood check-ins, stress levels, meditation logs, etc.), as well as monitor goal tracking and habit forming patterns – all extremely relevant in LM interventions. Evidence suggests that mHealth apps, web-based technologies, and telehealth technologies can improve chronic disease management, alleviate disease-related symptoms and patient adherence to interventions or medications.[18] Again, if the users give permission, their data can be anonymously shared with research platforms, healthcare professionals or coaches, or other institutes, to study the efficacy and impact of lifestyle interventions.
- **Electronic health records (EHRs):** Nowadays, health data from physicians, or other healthcare providers store data, such as medical history, symptoms, and diagnoses, clinical notes, prescriptions, treatment plans, and progress notes using EHRs. EHRs have also been widely adopted and evaluated on their accuracy to extract information.[19] Data is usually inputted manually either with free-text or voice dictation or through structured forms or checklists that the physicians will use to document vital signs, risk assessments, or screening questionnaires. EHRs can also be linked to laboratory systems (e.g., blood tests or imaging); medical devices (e.g., electrocardiogram monitors); wearable devices or mHealth technologies; as well as hospital systems that provide admissions, treatment, and discharge summaries. In relation to LM, EHRs allow clinicians and patients to set up specific goals or interventions and track progress at regular check-ins with the patient using blood biomarkers, imaging, or patient characteristics, such as weight and body mass index (BMI). With the development of AI and natural language processing to analyse free-text inputs,[20,21] EHRs are becoming more useable and connected and a reliable source of RWD on disease management.
- **Patient-reported outcomes (PROs):** PROs are a type of data collected often by clinicians and healthcare providers and are increasingly important in relation to LM as they help track interventions that are targeting a specific behaviour, quality of life, or the functional health of an individual. These are things only a patient can really report on. There are different methods to collect information – digitally, paper versions or verbally – depending on the outcome a physician or healthcare provider is assessing (Table 4).[22–37] PROs can be also be integrated into EHR systems for tracking, e.g., REDCap for clinical research, MyChart (a patient portal by Epic Systems), Apple Health, or Google Fit, as well as custom LM apps. Linking PROs with wearable devices is also possible by combining data collected with specific questions, such as "how did you sleep last night?".
- **Community health programmes:** Collecting meaningful, ethical, and actionable data using community health programmes can be a practical way to see how real-world settings paint a picture of health behaviour in a diverse population. Specific models, such as RE-AIM Framework exist, which help healthcare providers apply and assess community interventions and engagement (see Table 4 for links). Community health workers can conduct interviews or surveys (similar to the PROs mentioned above) and guide patients in using tools or devices to help gather quality data within a community. They can also provide health screening and pop-up clinics to regularly track blood pressure, waist

Table 4. Validated questionnaires to gather health data in lifestyle medicine

Lifestyle medicine pillar	Tool
Physical activity	IPAQ - International Physical Activity Questionnaire[22]
Nutrition	Rate Your Plate[23] DHQ – Diet History Questionnaire[24] MEDAS - Mediterranean Diet Adherence Screener[25]
Sleep	Pittsburgh Sleep Quality Index (PSQI)[26]
Stress / emotional health / mental health	Perceived Stress Scale (PSS)[27] Patient-reported outcomes measurement Information system (PROMIS)[28] Depression, Anxiety, and Stress Scale – 21 Items (DASS-21)[29] Patient Health Questionnaire-9 (PHQ-9) and PHQ-2[30]
Quality of life	RAND 36-item short form[31] Patient Activation Measure (PAM)[32]
Specific LM assessments	Lifestyle Medicine Assessment (LMA)[33] Hierarchies of Evidence Applied to LM (HEALM) assessment tool for studies[34]
Community models	RE-AIM (reach, effectiveness, adoption, implementation, maintenance)[35] PRECEDE-PROCEED Model[36] WHO STEPwise approach[37]

circumference, BMI, etc. and engage with patients and assist them to ensure they adhere to the intervention. Importantly, these programmes, such as nutrition classes, walking groups, etc. can gather data at baseline, during, or end of an intervention, at post-intervention, and at follow-up, e.g., 6–12 months after. These programmes can offer benefits with a more structured study design but are resource heavy.

What should we consider when using this data?

We need to be careful with any data, especially those that are not obtained from RCTs, i.e., obtained from studies not conducted in a controlled manner. Some key considerations for MWs when assisting researchers using RWD and publishing their findings would be:

- **Ethical consideration:** Ensure that informed consent was given for data collection either by the user accepting third-party sharing or sharing with care teams. Be transparent about the data and sources; providing storage links if necessary. Ensure the study is compliant with specific guidelines, such as GDPR or compliant with local health authorities and research ethics guidelines. Data should be protected, de-identified, and encrypted if digital, to ensure patient privacy.
- **Patient diversity and data variability:** RWD capture outcomes for diverse patient populations, including those with complications or vulnerabilities that are often excluded from RCTs.[34] However, they can also be restricted to a specific population and lack diversity, thus, findings would not be applicable to other populations. For example, many of the wearable devices and mHealth technologies are available in more middle- and high-income countries and therefore, lack data coming from low-income countries or regions with fewer resources. Consider the data source (EHR, app, registry, etc.), cultural context, socioeconomic status, and access to systems before generalising the findings and drawing a firm conclusion.
- **Long-term effects:** Lifestyle interventions typically require long study periods to show significant health impacts, which can be captured through RWD, but the data are often unreliable on the long-term effects of interventions due to data input, adherence, and patient participation. Consider the timeframes, the population group, and the inclusion and exclusion criteria, specifically examine the number of excluded individuals due to missing data points or drop-outs. Ideally, include a workflow to show data selection.
- **Comprehensive and quality of data:** RWD potentially provide large data sets that can reveal trends in various demographics and a more holistic view of patient health, including factors such as adherence, quality of life, and economic impacts. However, data quality and consistency is affected by the high variability in how people log their health activities and how accurately they do it, by study design (which is often poor and not adhered to across multiple sites), and by the bias that arises from self-reporting. Taken together, this means that it can be hard to draw meaningful conclusions. Check the study design and methodology, e.g., data source, unmeasured confounding factors, and consider the guidelines that were followed, as well as the statistical analyses that were done, e.g., propensity score matching and sensitivity analyses.

Reporting guidelines and checklists can help with this, e.g., STROBE, Hierarchies of Evidence Applied to LM (HEALM) (Table 4).[38]

Conclusion

The characteristics and trends around LM-related research and publications show that LM and its pillars are increasingly prominent. They are not isolated but rather permeate efforts to make progress in the management of multiple conditions. Additionally, the toolbox of methods to gather RWD has evolved enormously thanks to recent technological advances. Combining wearable data, mHealth technologies, EHRs, and PROs can create objectively measurable RWD of patient behaviour, disease progression, and treatment effectiveness. As these technologies and systems advance, so will RWD standards. This will drive LM's principles and interventions forward to combat the vast burden of NCDs. Thus, LM is definitely a medical field MWs of all specialisations should be aware of, as they could encounter opportunities within these areas or in studies and publications where LM constitutes part of the research. They should understand the characteristics of the body of evidence and the importance of RWD and RWE in driving the scientific evidence supporting this field.

Acknowledgements

The authors would like to thank Dr David Katz and Heather Mason for their insights.

Disclaimers

The opinions expressed in this article are the authors' own and not necessarily shared by their employers or EMWA.

Disclosures and conflicts of interest

The authors declare no conflicts of interest.

Data availability statement

The methods for the literature search appear in the Appendix following this article. For inquiries about data and other supplemental information, please contact the corresponding author.

References

1. Egger G, Binns A, Rössner S, et al. Editors. Lifestyle medicine: Lifestyle, the environment, and preventive medicine in health and disease. 3rd ed. Cambridge: Academic Press; 2017.
2. Frates B, Bonnet JP, Joseph R, et al. The lifestyle medicine handbook: An introduction to the power of healthy habits. 2nd. ed. Monterey: Healthy Learning; 2018.
3. Institute for Health Metrics and Evaluation (IHME). Global burden of disease 2021: Findings from the GBD 2021 study. 2024 [cited 2025 May 8]. Available from: https://www.healthdata.org/research-analysis/library/global-burden-disease-2021-findings-gbd-2021-study
4. WHO. Noncommunicable diseases. 2024 [cited 2025 May 8]. Available from: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases
5. Rippe JM, editor. Lifestyle medicine. 4th ed. Boca Raton: CRC Press; 2024.
6. Hâncu M, Hanes I, Mihălţan F. Lifestyle medicine and chronic diseases prevention and treatment. 2021 [cited 2025 May 8]. Available from: https://www.eulm.org/activities-education/research/elmo-e-book/
7. Miller WR. The evolution of motivational interviewing. Behav Cogn Psychother. 2023;51(6):616-32. doi:10.1017/S1352465822000431
8. Rojido MC. Catching the wave of lifestyle medicine. Med Writ. 2019;28(3):12-9.
9. Wilding JPH, Batterham RL, Calanna S, et al. Once-weekly semaglutide in adults with overweight or obesity. N Engl J Med. 2021;384:989-1002. doi:10.1056/NEJMoa2032183
10. Douglas MP, Kumar A. Analyzing precision medicine utilization with real-world data: A scoping review. J Pers Med. 2022;12(4):557. doi:10.3390/jpm12040557
11. Subbiah V. The next generation of evidence-based medicine. Nat Med. 2023;29(1):49-58. doi:10.1038/s41591-022-02160-z
12. Pirmani A, Moreau Y, Peeters LM. Unlocking the power of real-world data: A framework for sustainable he althcare. Stud Health Technol Inform. 2024;316:1582-3. doi:10.3233/SHTI240723
13. Ros E, Martínez-González MA, Estruch R, et al. Mediterranean diet and cardiovascular health: Teachings of the PREDIMED study. Adv Nutr. 2014;5(3):330S–6S. doi:10.3945/an.113.005389
14. Sherman RE, Anderson SA, Dal Pan GJ, et al. Real-world evidence – What is it and what can it tell us? N Engl J Med. 20168;375(23):2293-97. doi:10.1056/NEJMsb1609216
15. Anderson ML, Chiswell K, Peterson ED, et al. Compliance with results reporting at Clinicaltrials.gov N Engl J Med. 2015;372(11):1031-9. doi:10.1056/NEJMsa1409364
16. Hartung DM, Zarin DA, Guise JM, et al. Reporting discrepancies between the clinicaltrials.gov results database and peer reviewed publications. Ann Intern Med. 2014;160(7):477-83. doi:10.7326/M13-0480
17. Lu L, Zhang J, Xie Y, et al. Wearable health devices in health care: narrative systematic review. JMIR Mhealth Uhealth. 2020;8(11):e18907. doi:10.2196/18907
18. Fan K, Zhao Y. Mobile health technology: A novel tool in chronic disease management. Intell Med. 2022;2(1):41-7. doi:10.1016/j.imed.2021.06.003
19. Lossio-Ventura JA, Sun R, Boussard S, et al. Clinical concept recognition: Evaluation of existing systems on EHRs. Front Artif Intell. 2023;5:1051724. doi:10.3389/frai.2022.1051724
20. Al Ani M, Garas G, Hollingshead J, et al. Which electronic health record system should we use? A systematic review. Med Princ Pract. 2022;31(4):342-351. doi:10.1159/000525135
21. Kersloot MG, van Putten FJP, Abu-Hanna A, et al. Natural language processing algorithms for mapping clinical text fragments onto ontology concepts: a systematic review and recommendations for future studies. J Biomed Semantics. 2020;11(1):14. doi:10.1186/s13326-020-00231-z
22. International Physical Activity Questionnaire: IPAQ scoring protocol. IPAQ Group [cited 2025 Jul 7]. Available from: https://sites.google.com/view/ipaq/download
23. UCLA Health. Rate Your plate: Heart. 2010 [cited 2025 Jul 7]. Available from: https://www.uclahealth.org/sites/default/files/documents/Rate%20Your%20Plate%20Heart%202010.pdf
24. National Cancer Institute. Diet History Questionnaire III (DHQ III). n.d. [cited 2025 Jul 7]. Available from:

https://epi.grants.cancer.gov/dhq3/
25. Hamilton Health Sciences. Mediterranean Diet Adherence Screener (MEDAS). 2017 [cited 2025 Jul 7]. Available from: https://www.hamiltonhealthsciences.ca/wp-content/uploads/2022/04/Cardiac-Health-Mediterranean-Diet-Adherence-Screener.pdf
26. University of Pittsburgh. Pittsburgh Sleep Quality Index (PSQI). n.d. [cited 2025 Jul 7]. Available from: https://www.sleep.pitt.edu/psqi
27. NovoPsych. Perceived Stress Scale (PSS-10). n.d. [cited 2025 Jul 7]. Available at: https://novopsych.com/assessments/well-being/perceived-stress-scale-pss-10/
28. National Institutes of Health. Patient-reported outcomes measurement information system (PROMIS). 2025 [cited 2025 Jul 7]. Available from: https://commonfund.nih.gov/patient-reported-outcomes-measurement-information-system-promise
29. Motor Accident Insurance Commission. Depression anxiety stress scales – DASS-21. n.d. [cited 2025 Jul 7]. Available from: https://maic.qld.gov.au/wp-content/uploads/2016/07/DASS-21.pdf
30. American Psychological Association. Patient Health Questionnaire (PHQ-9). 2020 [cited 2025 Jul 7]. Available from: https://www.apa.org/pi/about/publications/caregivers/practice-settings/assessment/tools/Patient-health
31. RAND Corporation. 36-Item Short Form Survey (SF-36). n.d. [cited 2025 Jul 7]. Available from: https://www.rand.org/health-care/surveys_tools/mos/36-item-short-form.html
32. Hibbard J H, Stockard J, Mahoney ER, et al. Development of the patient activation measure (PAM): conceptualizing and measuring activation in patients and consumers. Health Serv Res. 2004;39 (4 Pt 1):1005-26. doi:10.1111/j.1475-6773.2004.00269.x
33. Fundación MA. Survey on health and lifestyle habits. n.d. [cited 2025 Jul 7]. Available from: https://survey.fundacionmapfre.org/index.php/211124?lang=en
34. Dreyer NA. Strengthening evidence-based medicine with real-world evidence. Lancet Healthy Longev. 2022;3(10):e641-2. doi:10.1016/S2666-7568(22)00214-8
35. RE-AIM. Resources and tools. n.d. [cited 2025 Jul 7]. Available from: https://re-aim.org/resources-and-tools/
36. Rural Health Information Hub. Precede-proceed model. 2024 [cited 2025 Jul 7]. Available from: https://www.ruralhealthinfo.org/toolkits/health-promotion/2/theories-and-models/precede-proceed
37. WHO. STEPwise approach to NCD risk factor surveillance (STEPS). n.d. [cited 2025 Jul 7]. Available from: https://www.who.int/teams/noncommunicable-diseases/surveillance/systems-tools/steps
38. Katz DL, Karlsen MC, Chung M, et al. Hierarchies of evidence applied to lifestyle medicine (HEALM): Introduction of a strength-of-evidence approach based on a methodological systematic review. BMC Med Res Methodol. 2019;19(1):178. doi:10.1186/s12874-019-0811-z

Author information

Maria Carolina Rojido is a physician with a master's degree in health administration and certificates in lifestyle medicine and nutrition coaching. She has extensive clinical research and medical writing experience (9 and 12 years) in multiple diseases, treatment approaches, and deliverables. Her nutrition and lifestyle medicine journey started 10 years ago.

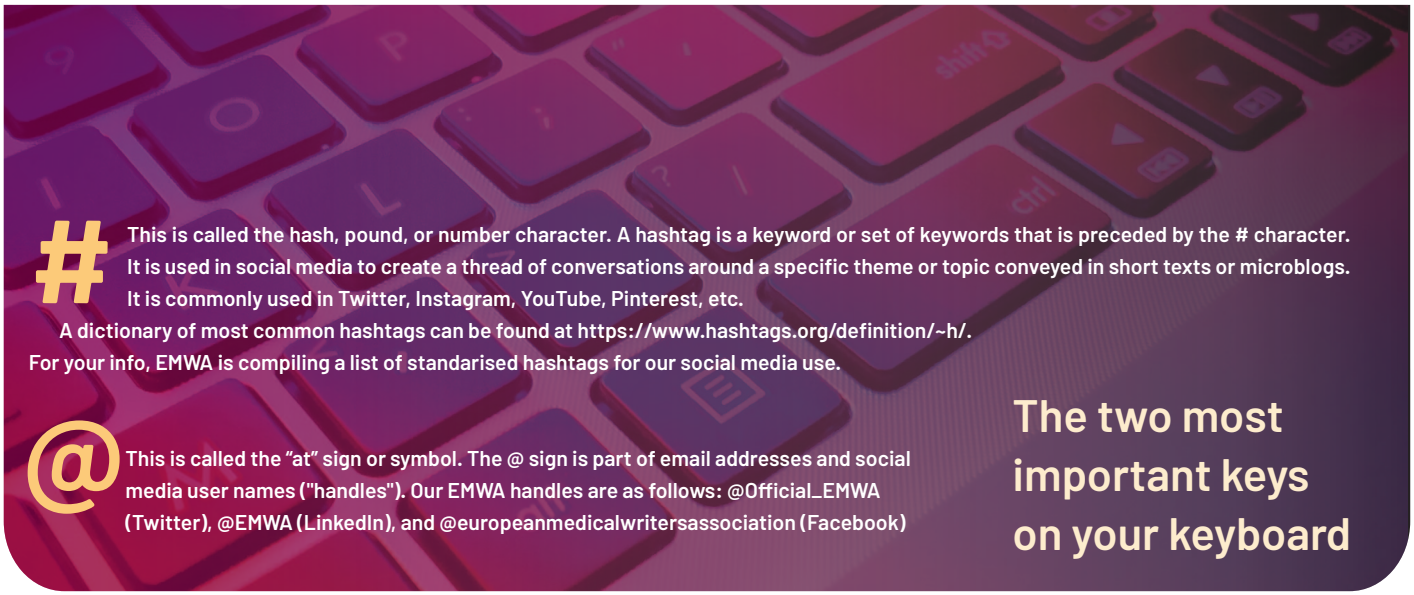
Mariana Rickmann is a biochemist with a PhD in biomedicine, and a lifestyle medicine consultant, certified by ELMO. She has been working as a freelance medical writer since 2017, translating science for the public and collaborating with startups and media such as In a nutshell-Kurzgesagt. She's currently working on several projects around lifestyle medicine.
Google scholar profile:
https://scholar.google.de/citations?hl=es&user=aWlcPhcAAAAJ

Laura A. Kehoe holds a PhD in neurobiology and is a freelance medical communications specialist with extensive experience in publications, grant funding and online communications. She is also a qualified Integral Coach® and certified lifestyle medicine advisor specialising in women's health and neurodiversity.
https://www.linkedin.com/in/laura-a-kehoe-phd-93893982/

Appendix: Methodology of clinical trials and publications searches on lifestyle medicine

To find out the types of studies in LM, we searched in clinicaltrial.gov and PubMed. First, we performed a search on clinical trials.gov for observational and interventional studies where the intervention was lifestyle (this was the exact term used); no other filters were applied. We excluded those with suspended, unknown, or withdrawn study status. From the resulting list, we used the "sort" function in Excel and searched on the study titles and interventions columns using text filters for words representative of LM and its key methodology: lifestyle, coaching, health coaching, counselling, motivational, habits, and behaviour. We also searched terms related to the eight pillars of LM in the study title and intervention columns. We focused on four pillars: diet, physical activity, stress, and sleep because they contained much more entries than the other pillars. Thus, the rows with the following representative words were included: nutrition and diet; exercise and physical activity; sleep; stress, mindfulness, and meditation. We also searched the location column with text filters for rows containing country names and counted those with >10 studies. We ordered the study design column and obtained the numbers and subtypes of randomised studies. Lastly, we ordered the start year column and counted the number of studies that fell into 5-year subcategories. We also counted and classified the number of studies with more than one intervention. Lastly, we colour-coded, and made an initial count of the most common conditions treated.

Secondly, to find out the types of publications in our areas of interest, we performed targeted searches in PubMed with the following search strings: lifestyle[Title]) AND (1995:2025 [pdat]), ((nutrition[Title]) OR (diet[Title]) OR (dietary[Title]) OR (food[Title])) AND (1995: 2025[pdat]), ((stress[Title]) OR (meditation [Title]) OR (mindfulness[Title])) AND (1995: 2025[pdat]), ((exercise[Title]) OR (physical activity[Title])) AND (1995:2025 [pdat]), ((sleep[Title])) AND (1995:2025 [pdat]). We then used the filters for the following article types: observational and case reports; clinical trial and randomised clinical trial; systematic reviews and meta-analyses; reviews; practice guidelines and guidelines; editorials; and letters.




Barcelona
SPAIN
May 5–8, 2026
EMWA's 61st Conference
SAVE THE DATE

Silenced data: How banning words undermines real-world evidence in medical writing

Giovanna Gelmi[1], **Claudia Percivalle**[2]
Freelance medical writers
[1] Aachen, Germany
[2] Milano, Italy

doi: 10.56012/xgwp6546

Correspondence to:
Giovanna Gelmi
giovanna@giovannagelmi.de
Claudia Percivalle
claudia.percivalle@sciencewriting.it

Abstract
Executive orders (EOs) issued by the President of the United States can significantly shift federal research priorities, funding allocations, and public health directives, thereby influencing which medical topics receive attention and resources. EOs also affect the transparency, availability, and regulation of medical data. In this article, we report how language censorship brought about by recent EOs affects the collection, interpretation, and communication of real-world evidence. Real-world evidence depends on accurate, inclusive, and standardised terminology. Banning certain words undermines data integrity and scientific utility.

Image: Freepik

US executive actions on real-world evidence, 2016–2025

This year has witnessed a revival of what had already happened to a lesser extent in 2017, that is, the disappearance of certain words from scientific documents and official government websites in the United States, but this time it has occurred with much greater intensity. These are the so-called "word bans" that followed the executive orders of the White House.[1,2] However, the White House denied the existence of a list of prohibited words.[3] An official Executive Order (EO) banning specific words does not, in fact, exist.[4,5] To grasp what happened, we must first understand the EO mechanism. An EO is an official act issued by the President of the United States. Although these orders are not laws, they are a primary tool by which the President can direct the operations of the federal government.

These policies have shaped the reporting of clinical and epidemiological information – including real-world evidence (RWE), defined by FDA as clinical evidence about the use and benefits/risks of medical products derived from analyses of real-world data (such as electronic health records, insurance claims, and patient registries).[6] The US executive decisions, ranging from memoranda to EOs, from 2016 to date, that have a significant impact on RWE are shown in Figure 1 and described in Table 1.

> Although these orders are not laws, they are a primary tool by which the President can direct the operations of the federal government.

The "banned" words

In March 2025, *The New York Times*,[7] based on publicly available texts of the EOs published in the US Federal Register, compiled a list of 197 words or concepts that agencies had flagged to limit or avoid, resulting from EOs issued this year.[8] The list, available at the *New York Times* website, starts with "accessible", ends with "women and underrepresented", and includes many terms related to diversity, equity, and inclusion (DEI), as well as climate science. These words and phrases were being removed from websites and replaced with others deemed acceptable by the current administration. *The New York Times* also provided examples of how words had been deleted, such as the visual depiction of changes to a memo about Head Start, a US programme to

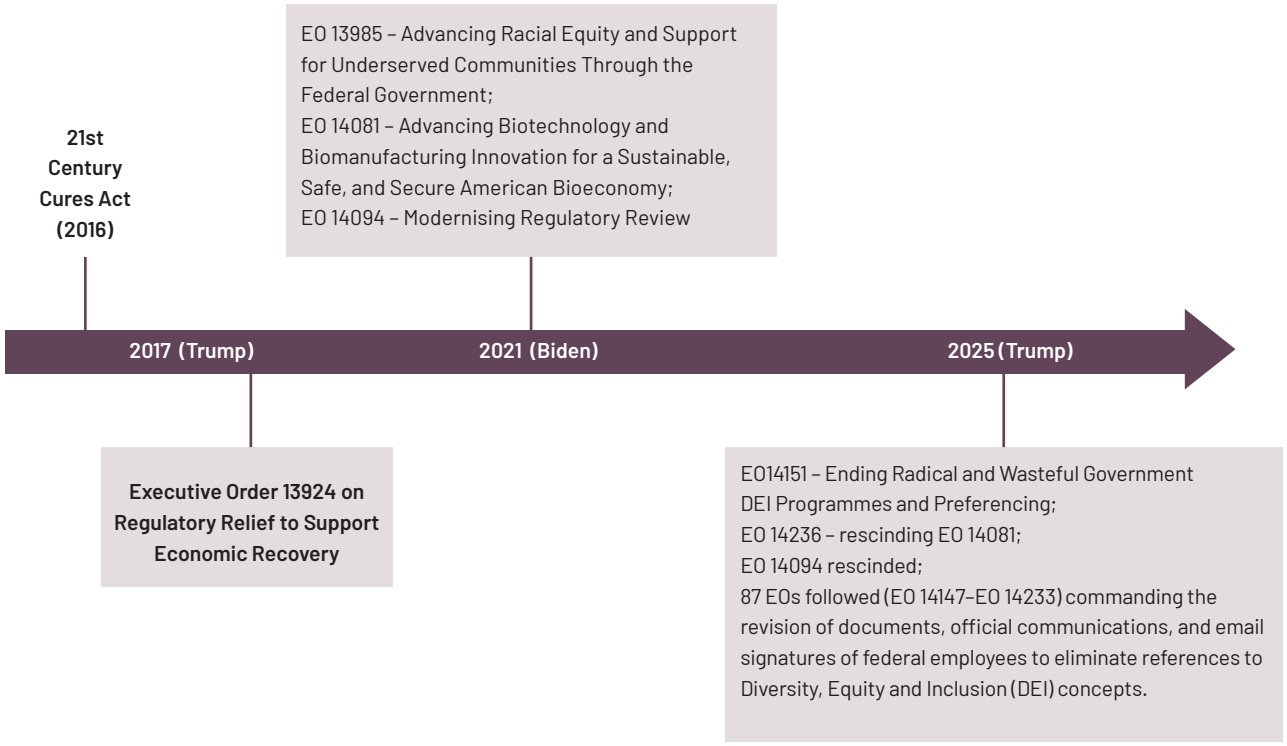


Figure 1. Figure that reproduces with a timeline the different Executive Orders and Memorandum during Biden and Trump Administrations around DEI

promote early childhood education for children in lower-income families:

> The last year has brought significant challenges to the Head Start workforce. ~~The COVID-19 pandemic has had a disparate impact on under-resourced communities including many of those served by Head Start programmes.There has also been heightened attention to racial injustice in our country, which has led to calls for major reforms to address long-standing societal inequities. These are particularly important concerns for OHS and the Head Start workforce.~~ All staff have been impacted by COVID-19. ~~Further, 60% of Head Start teaching staff are Black, Indigenous and people of colour, and 30% have a primary language other than English.~~ As such, OHS is committed to a culture of wellness that includes holistic support for the entire Head Start workforce.

Darby Saxbe,[9] a professor at the University of Southern California, posted on social media an example of how specific uses of language were being reviewed to determine which health grants should be canceled (Figure 2). The decision tree was sent to, among others, all programme officers at the National Science Foundation (NSF).[10]

As as a result of the White House EOs, operators of individual agencies were tasked with deciding whether a term should be removed, replaced, or retained, depending on the context. In addition to their hierarchical administrative organisation, the agencies of the federal government of the United States are interconnected at multiple levels,[11] through hyperlink and datalink paths across the web and linked open data (LOD). Therefore, changes in the semantics of any one of the sites with the .gov extension can indirectly influence the interpretation or use of terms in other .gov-linked sites, especially where there is semantic overlap or hyperlink-based data referencing. Changes in one site do not automatically update others, but they can cause misalignment, misunderstanding, or require re-interpretation downstream. Thus, there are many opportunities for inconsistent or contradictory uses of terminology and phraseology, across governmental agencies and contexts.

One of the affected datasets is CDC's Behavioral Risk Factor Surveillance System (BRFSS),[12] which is one of the most widely used national health surveys and has been ongoing for about 40 years. BRFSS has been used for decades to inform policymakers, the media, and the public on a wide range of health topics, such as obesity rates, access to breast cancer screenings, vaccination rates, and the proportion of people with pre-existing conditions. With sampling in every state, BRFSS data are particularly helpful for understanding health issues in low-population states and rural areas. It was briefly taken offline and later returned without its questionnaires or codebooks. However, without that documentation, researchers cannot verify how variables were measured or replicate analyses, undermining the integrity of any RWE derived from those data.

In total, roughly 8000 federal web pages disappeared from public view (some later returned with warning banners like "CDC's website is being modified to comply with President Trump's Executive Orders") but some

Year	EO / Memo	Description
2016	**21st Century Cures Act**	Mandated the FDA to evaluate how RWE can support approval of new indications for approved drugs and post-approval study requirements. A major legislative foundation for RWE.
2020	**Executive Order 13924 on Regulatory Relief to Support Economic Recovery**	Prioritised deregulation and reduction of data/reporting burdens, which may have limited RWE infrastructure development.
2021	**EO 13985 – Advancing Racial Equity and Support for Underserved Communities Through the Federal Government**	Promotes equity in data collection and health research, which enables inclusive RWE generation and use.
2021	**Memo: "Restoring Trust in Government Through Scientific Integrity and Evidence-Based Policymaking."**	Federal policy must be "guided by the best available science and data" and "scientific findings should never be distorted or influenced by political considerations."
2021–2024	**FDA RWE Guidance Series (e.g., on data standards, study design, and regulatory use)**	Though not EOs, these guidance documents support and operationaliSe the RWE programme under the Cures Act.
2022	**EO 14081 – Advancing Biotechnology and Biomanufacturing Innovation for a Sustainable, Safe, and Secure American Bioeconomy**	Encourages data innovation and evidence generation, including the use of RWE for regulatory and clinical applications.
2024	**EO 14094 – Modernising Regulatory Review**	Promotes evidence-based decision-making, encouraging agencies to use modern data approaches, potentially including RWE.
2025	**EO 14151 – Ending Radical and Wasteful Government DEI Programs and Preferencing**	Targets diversity programmes that are essential to equitable RWE generation; may roll back inclusive data strategies.
2025	**EO 14236, rescinding EO 14081. EO 14094 rescinded**	Deregulatory moves reducing support for RWE, particularly those rooted in DEI, data modernisation, or government health innovation.

Table 1. Executive orders and memoranda with implications for real-world evidence

President Trump's first term was from January 2017 to January 2021, then he was returned to office in January 2025. President Biden served in the 4 years in between.

Box 1. Partial list of US federal health data that had been taken offline at least temporarily

US Centers for Disease Control and Prevention (CDC): AtlasPlus; an interactive database with about 15 years of surveillance data for HIV, viral hepatitis, sexually transmitted diseases, and tuberculosis, as well as data on the social determinants of health.

PEPFAR Data Dashboards: PEPFAR, the US global HIV/AIDS Programme, comprehensive, up-to-date online data portal of program budgets and expenditures by country and service category.

Demographic and Health Surveys (DHS) databases: Data downloads from the DHS, an ongoing set of nationally representative household surveys supported by USAID, the US, international development agency, with population, health, HIV, and nutrition data from more than 90 countries.

Foreignassistance.gov: The US government's website with all foreign assistance data by country, budget, expenditure, programme

Area Health Resource Files: a resource of data on health professionals, hospitals, and economic CDC's Social Vulnerability Index: Census-based socioeconomic data used for disaster planning, response and recovery

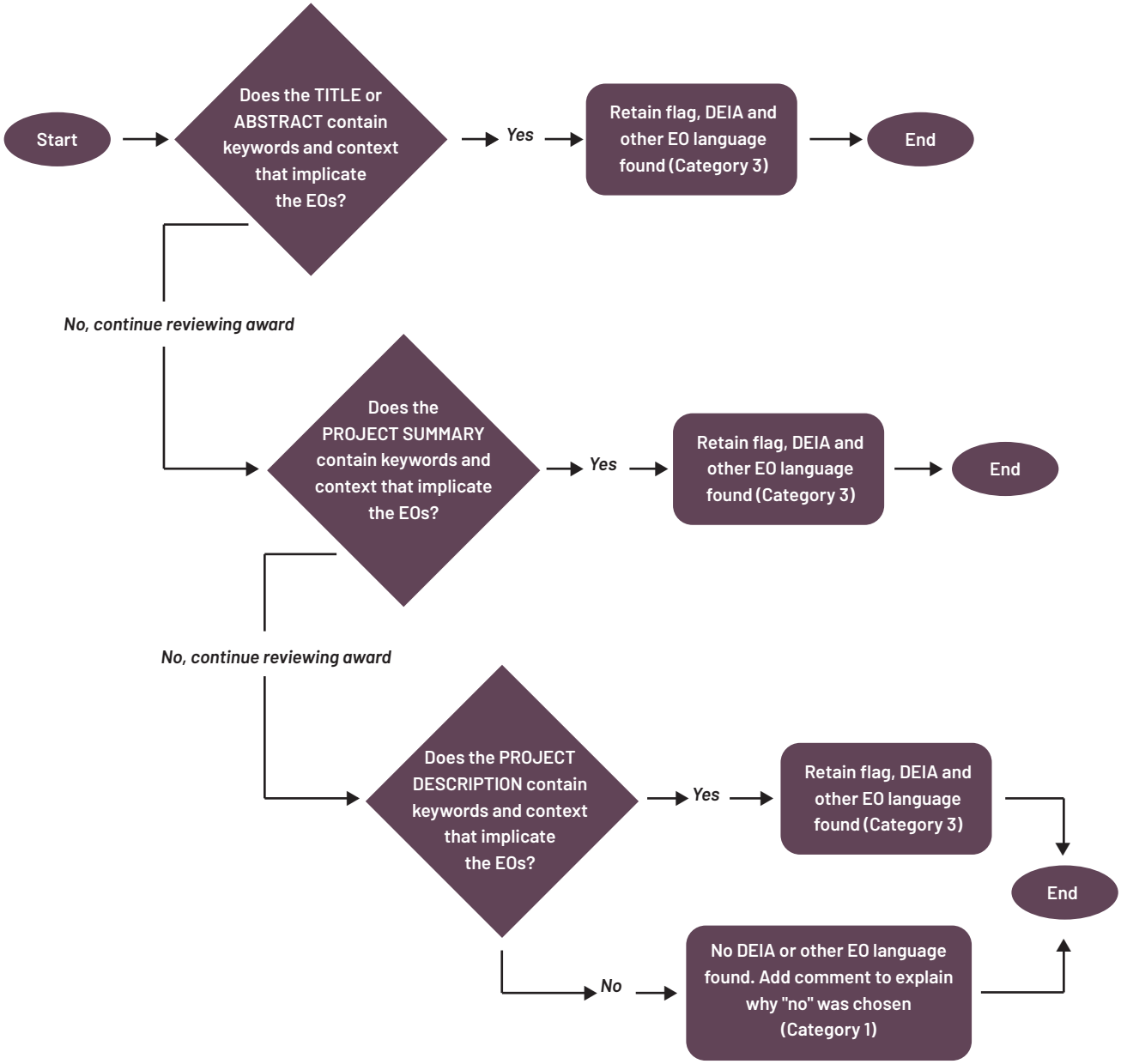


Figure 2. A decision tree distributed to program officers at the National Science Foundation to consider whether certain grants should be cancelled to comply with policies of the Trump Administration

A university professor posted the original image on social media. It was updated by the journal to improve clarity of the low-resolution image.

crucial websites are still not available (e.g., https://reproductiverights.gov/).[13] A list of federal health data sites that were at least temporarily taken offline and/or later altered is provided in Box 1.

"Bias" as a banned word

The term *bias,* far from being a hallmark for DEI topics only, is a foundational concept in knowledge and science. The phrase "cognitive bias" was introduced in the early 1970s by psychologists Amos Tversky and Daniel Kahneman to indicate systematically flawed patterns of responses to judgment and decision problems.[1] In 2002, Kahneman was awarded the Nobel Prize in Economics with the motivation "for having integrated insights from psychological research into economic science, especially concerning human judgment and decision-making under uncertainty".[15]

In medicine, too, human judgement and decision-making under uncertainty play pivotal roles – by patients, physicians, healthcare professionals, or scientists. Indeed, an increasing number of cognitive biases, from framing to anchoring to status-quo bias, have been recognised in medical science and practice over the last decades.[16]

A PubMed search using the terms *bias* and *human research*, thus excluding animal and pure laboratory research, yielded over 68,000 results from years 1966–2025, over 65,000 of them from years 2000–2025. The medical community at large is now aware that our attempts to understand reality are flawed, i.e., biased, and accounts for those biases, routinely

implementing corrective measures. Banning the word *bias* equates to sabotaging efforts to understand reality as it is, and RWE as its most appropriate measure.

How RWE is related to terminology

Three key regulatory elements must be in place for RWE to be effective: RWE regulatory framework, data quality and standards guidance, and study methods guidance. We focus here on the second element: data must be available, accessible, and fit for use. And, possibly, even improved upon: initiatives for ensuring high-quality RWD availability, access, standardisation, and methodological rigour have been advocated in pursuit of ever higher-quality RWE. This is even more true now in the era of big data. The Big Data Task Force was created in 2017 jointly by EMA and HMA (Heads of Medicines Agencies) to tackle the challenges posed and reap the opportunities offered by big data.[17]

Data standardisation relies on terminology, defined by the NIH as "a systematically organised set of terms, concepts, and codes used in health care to describe clinical conditions, procedures, medications, and other healthcare-related topics in a consistent and uniform manner, while a term is defined as "human readable text description that can act as the anchor meaning for the concept".[18] So, though a term does not equate with a concept, the two are inextricably bound. The loss of a term starves the concept anchored to that term and, vice-versa, the free usage of a term is instrumental for the anchored concept to be circulated and elaborated on.

Patient demographics

In medicine, baseline data on demographics are the important starting point, including data on gender, race, and ethnicity. In March 2024, the Federal Register published the Standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity[19] to improve the quality and usefulness of federal race and ethnicity data. The document recommends that information on race and ethnicity be collected using a single question that combines both, moving from two separate questions. This comes as a consequence that "since 1980, responses to the decennial census in each subsequent decade have shown increasing non-response to the race question, confusion, and concern from the public about separate questions on ethnicity and race". The Standards now define seven race or ethnic groups all of them to be used alone or in combination according to three different Approaches, plus the newly introduced "multi-racial and/or multiethnic group" introduced in Approach 3.

The updates, therefore, try to reflect the current multifaceted reality to the best of their capabilities. They are inspired by the idea that templates should reflect reality, not reality be moulded to adjust to templates. Similarly, the NIH directs the use of Office of Management and Budget (OMB) census categories with self-identification for clinical trials, to make such artificial settings, as clinical trials are, as realistic as possible. Any removal of the terms indicating ethnic groups as commanded by the US administration would yield data that do not accurately describe reality. What is more, the removal of the very terms "race" and "ethnicity", as stated in the word bans, implies that it is deemed that neither concept has any relevance in medicine. This is known to be untrue for either genetic, environmental, social or cultural reasons, often for some or all of them combined.

The word bans will deprive PRO materials developed in the US of a wide range of commonly used terms, which are meaningful and unequivocal to patients and health care providers alike.

Medical outcomes

Another key set of data in medicine are outcome data. Medical outcomes in general can differ due to variations in drug pharmacokinetics or pharmacodynamics, or both, based on different age, race, and ethnic groups, clinical and other conditions, as well as genetic variants and gender. In particular, evidence of drug effects differing by gender has been documented for a long time in both clinical trials and real-life settings. A UK general practice study, combining 48 national cohort studies of newly marketed drugs, and comprising over 500,000 patients, reported that suspected adverse drug reactions to drugs are 60% more common in women than in men.[20] Drug gender differences exist in effectiveness, too. Low-dose aspirin tested on almost 40,000 patients has no significant efficacy on the risk of myocardial infarction or death from cardio-vascular causes in women, as opposed to results in men.[21] It was proven that dosing, too, can require massive adjustments in women.[22] Given the broad range of proven gender differences in drug effects, the amount of data available from both clinical trials and real-life practice, and the long time for which such knowledge has been around, gender stands out as a parameter that cannot be overlooked in medicine at any stage. A ban on the words "woman, women" would make it impossible to present data by gender, thus completely failing to reflect reality for either men or women.

Patient-reported outcomes

Finally, word bans would affect situations that we have come to realise more recently. Although the English language holds a global standing and is often the source language for translation, the vast majority of patients worldwide routinely receive and provide medical information in their own language, i.e., a language other than English. This fact will remain for the future, dictated by reality and mandated by national legislations. The last few years have seen a considerable effort in linguistic validation of medical translations which directly or indirectly target patients.

In particular, linguistic validation (LV) of patient-reported outcomes (PROs), such as questionnaires and rating scales, is a critical component in modern clinical research and, increasingly, in real-world studie.[23] Linguistic validation is a process that ensures that translated content accurately represents the source while being culturally and linguistically appropriate for the target population. LV ensures that PRO instruments maintain linguistic accuracy, cultural relevance, and conceptual equivalence to the original version. The process involves a) forward translation and back translation to preserve meaning; b) cognitive debriefing with target-language patients to validate comprehension; c) in the case of multinational trials, regulatory alignment with the FDA, EMA, and other agencies that require proof of equivalence.[24]

The word bans will deprive PRO materials developed in the US of a wide range of commonly used terms, which are meaningful and unequivocal to patients and health care providers (HCPs) alike, thus rendering source texts less

Image: Freepik

comprehensible to patients, and translations either non-viable or invalid. The number of viable source texts for PRO translations will drop, and this will impact patients in real-life practice worldwide.

Reactions in defence of RWE integrity

There have been reactions from US scientists aimed at preserving the integrity of real-world evidence (RWE), including the reversal of language bans, the republication of vital datasets, and the reaffirmation of evidence-based standards in agency guidelines.

Some concrete initiatives to defend RWE integrity are as follows:

- Scientists, advocates, and institutions are mobilising to protect data and defend the principles of evidence-based research. To save federal health websites and databases, researchers are using different tools, including downloading datasets, scraping websites and archiving them with the Wayback Machine,[25] which is an initiative of the Internet Archive, and enables users to see how websites looked in the past.
- The Association of Health Care Journalists protested the removal of public health data "at a time when the rise in chronic illnesses and harmful behaviors among young people is at the top of the national agenda".[26]
- The American Medical Writing Assocation reacted by reaffirming its values and mission relating to DEI in a message to members.[27]

In the rest of the world, scientists and researchers are showing solidarity with their US colleagues. Here are a few examples:

- A coordinated stand by international publishers (ICMJE editors) defending evidence-based standards is the commentary in Lancet (co-signed by editors around the world) explicitly denounced the US policies as "part of a global assault on evidence, inclusion, and truth," urging that scientists, publishers and editors "must resist silence" in the face of censorship.[28]
- The nonprofit publisher PLOS (USA/global) issued a forceful blog statement reaffirming its commitment to open, rigorous science.[29]
- In *Nature Medicine*, van Daal et al. explicitly warned that banning words in medical research is "bad news for everyone".[30]
- Other countries' journals and experts have echoed these concerns. For example, an editorial in Tobacco Control (Australia) warned that the new U.S. administration has enacted "savage cuts to health research, agencies and programmes; attempts to prevent, retract or amend scientific publications; [and] deletion of health databases".[31]

Conclusions

Terminology accuracy is essential to provide understandable and meaningful RWE information. Scientists and writers should be free to use all terms that have been developed across disciplines over time and have been demonstrated to be sound and valid for their intended purposes. The loss of that accuracy or the elimination of context-specific terms can deprive decision-makers of vital information.

The recent and continuing censorship policies described this article underlines

- the political vulnerability of health data systems and the implications on global research reliability.
- the need for international standards for data governance and independent audits.

As medical writers and communicators, we are aware that RWE depends fundamentally on the availability, transparency, and integrity of large-scale health data – domains in which the U.S. has historically been a global leader. However, if data is selectively removed, censored, or altered for ideological or political purposes, the very reliability of RWE as a scientific tool is called into question. This not only affects the credibility of US-based data sources but also the trustworthiness of any evidence derived from them. The medical writing community can contribute to safeguarding the ethical use of RWE by building international standards for data governance and independent audits.

Disclaimers

The opinions expressed in this article are the authors' own and not necessarily shared by EMWA. ChatGPT Deep Search was used as a reference tool; each single reference included was verified independently.

Disclosures and Conflicts of Interests

The authors declare no conflict of interest.

Image: Freepik

References

1. Mahdawi A. Saying 'women' is not allowed, but 'men' and 'white' are OK? I'm (not) shocked. The Guardian; 2025 Feb 15 [cited 2025 May 28]. Available from: https://www.theguardian.com/commentisfree/2025/feb/15/trump-banning-words-women
2. Amon JJ. Trump's banned words and disastrous health policies. HHR 2025;27(1):83-6. Available from: https://www.hhrjournal.org/2025/02/04/trumps-banned-words-and-disastrous-health-policies/
3. Levy R. Exclusive: FDA staffers told that 'woman,' 'disabled' among banned words; White House says it's an error. Reuters; 2025 Feb 20 [cited 2025 May 28]. Available from: https://www.reuters.com/business/healthcare-pharmaceuticals/fda-staffers-told-that-woman-disabled-among-banned-words-white-house-says-its-an-2025-02-20/
4. Executive Orders. The White House. [cited 2025 May 28]. Available from: https://www.whitehouse.gov/presidential-actions/executive-orders/
5. Executive Orders. Federal Register[cited 2025 May 28]. Available from: https://www.federalregister.gov/presidential-documents/executive-orders
6. Real-world evidence. U.S FDA [cited 2025 May 28]. Available from: https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence
7. Yourish K, Daniel A, Datar S, et al. These words are disappearing in the new Trump Administration The New York Times; 2025 Mar 7 [cited 2025 May 28]. Available from: https://www.nytimes.com/interactive/2025/03/07/us/trump-federal-agencies-websites-words-dei.html
8. Federal Register. The Daily Journal of the United States Government [cited 2025 May 28]. Available from: https://www.federalregister.gov/
9. USC Dornsife College of Letters, Arts and Science. Darby Saxbe, Professor of Psychology [cited 2025 May 28]. Available from: https://dornsife.usc.edu/profile/darby-saxbe/
10. Novak M. The list of Trump's forbidden words that will get your paper flagged at NSF. Gizmodo.com; 2025 Feb 5 [cited 2025 May 28]. Available from: https://gizmodo.com/the-list-of-trumps-forbidden-words-that-will-get-your-paper-flagged-at-nsf-2000559661
11. Norwich University. Intergovernmental relationships: How the three levels of government work together [cited 2025 May 28]. Available from: https://online.norwich.edu/online/about/resource-library/intergovernmental-relationships-how-three-levels-government-work-together
12. CDC Behavioral Risk Factor Surveillance System [cited 2025 May 28]. Available from: https://www.cdc.gov/brfss/index.html
13. Feldscher K. As health data disappear from government websites, experts push back. Harvard T.H. Chan School of Public Health; 2025 Feb 5 [cited 2025 May 28]. Available from: https://hsph.harvard.edu/news/as-health-data-disappear-from-government-websites-experts-push-back/
14. Kahneman D. A perspective on judgment and choice: mapping bounded rationality. Am Psychol. 2003;58(9):697-720. doi:10.1037/0003-066X.58.9.697
15. Princeton University. News. Daniel Kahneman, pioneering behavioral psychologist, Nobel laureate and giant in the field, dies at 90cited 2025 May 28]. Available from: https://www.princeton.edu/news/2024/03/28/daniel-kahneman-pioneering-behavioral-psychologist-nobel-laureate-and-giant-field
16. Hammond MEH, Stehlik J, Drakos SG, et al. Bias in medicine: Lessons learned and mitigation strategies. JACC Basic Transl Sci. 2021;6(1):78-85. doi:10.1016/j.jacbts.2020.07.012

17. Burns L, Roux NL, Kalesnik-Orszulak R, et al. Real-world evidence for regulatory decision-making: Guidance from around the world. Clin.Ther. 2022;44(3):420-37. doi:10.1016/j.clinthera.2022.01.012
18. NIH. Health Data Standards and Terminologies [cited 2025 May 28]. Available from: https://www.nlm.nih.gov/oet/ed/healthdatastandards/02-300.html#:~:text=A%20standardized%20medical%20terminology%20is,a%20consistent%20and%20uniform%20manner.
19. US Federal Register. Revisions to OMB's Statistical Policy Directive No. 15: Standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity [cited 2025 May 28]. Available from: https://www.federalregister.gov/documents/2024/03/29/2024-06469/revisions-to-ombs-statistical-policy-directive-no-15-standards-for-maintaining-collecting-and
20. Martin RM, Biswas PN, Freemantle SN, et al. Age and sex distribution of suspected adverse drug reactions to newly marketed drugs in general practice in England: analysis of 48 cohort studies. Br J Clin Pharmacol. 1998;46(5):505-11. doi:10.1046/j.1365-2125.1998.00817.x
21. Ridker PM, Cook NR, Lee IM, et al. A randomized trial of low-dose aspirin in the primary prevention of cardiovascular disease in women. N Engl J Med. 2005;352:1293-1304. doi:10.1056/NEJMoa050613
22. Farkas RH, Unger EF, Temple R. Zolpidem and driving impairment—identifying persons at risk. N Engl J Med. 2013;369(8):689-91. doi:10.1056/NEJMp1307972
23. McKown S, Acquadro C, Anfray C, et al. Good practices for the translation, cultural adaptation, and linguistic validation of clinician-reported outcome, observer-reported outcome, and performance ou tcome measures. J Patient Rep Outcomes. 2020;4(1):89. doi:10.1186/s41687-020-00248-z
24. Wild D, Eremenco S, Mear Iet al. Multinational trials-recommendations on the translations required, approaches to using the same language in different countries, and the approaches to support pooling the data: the ISPOR Patient-Reported Outcomes Translation and Linguistic Validation Good Research Practices Task Force report. Value Health. 2009;12(4):430-40. doi:10.1111/j.1524-4733.2008.00471.x
25. Internet Archive [cited 2025 May 28]. Available from: https://web.archive.org/
26. Freyer FJ, Gil G, Waner A. An urgent message from the nation's health journalists. Association of Health Care Journalists 2025 Jan 31 [cited 2025 Jul 21]. Available from: https://healthjournalism.org/blog/2025/01/an-urgent-message-from-the-nations-health-journalists/
27. A note from AMWA leadership. American Medical Writers Associations. 2025 Feb 14 [cited 2025 Jul 21]. Available from: https://www.amwa.org/news/693817/A-Note-from-AMWA-Leadership.htm?utm_content=324844369&utm_medium=social&utm_source=linkedin&hss_channel=lcp-659269
28. Frizelle F, Abbasi K, Bachelet VC, et al. Global cost of silencing science: editors and publishers have a duty to resist. Lancet. 2025;406(10505)789–91. doi:10.1016/s0140-6736(25)01429-1
29. PLOS statement on recent US Executive Orders and scientific integrity. The Official PLOS Blog. 2025 Feb 21 [cited 2025 May 28]. Available from: https://plos.io/3D4O8cH
30. van Daal M, Milota M, Jongsma KR. Why banning words in medical research is bad news for everyone. Nat Med. 2025;31(5):1388-1389. doi:10.1038/s41591-025-03596-9
31. Hefler M. Responding to US executive orders and Trump 2.0: ideological language bans and attempts to suppress science will not obscure the tobacco epidemic's foundations of exploitation and inequity. Tob Control. 2025;34:141–2. doi:10.1136/tc-2025-059437

Author Information

Giovanna Gelmi works as a freelancer in food and pharmaceutical industries. Her academic background includes two master's degrees (Food Science and Agricultural Sciences) along with a master's in project management. She previously worked as a clinical data manager.

Claudia Percivalle brings 15 years of experience in clinical research, R&D project management, and regulatory affairs, supported by a degree in pharmaceutical chemistry. Her background as a medical translator has deepened her passion for language and sharpened her ability to communicate clearly, always with the reader and listener in mind.

News from the EMA

The articles included in this section are a selection from the European Medicines Agency (EMA)'s News and Press Releases archive. More information can be found on the Agency's website: www.ema.europa.eu.

SECTION EDITOR

Section Editor:
Anuradha Alahari
Anuradha.Alahari@parexel.com

EMA contact:
Monika Benstetter
press@ema.europa.eu

Image: Freepik

New treatment for adults with acute lymphoblastic leukaemia: Third CAR T-cell therapy for high-mortality cancer

May 23, 2025

EMA has recommended granting a conditional marketing authorisation in the European Union (EU) for Aucatzyl (obecabtagene autoleucel) to treat adults from 26 years of age with relapsed or refractory B-cell precursor acute lymphoblastic leukaemia (B ALL).

Acute lymphoblastic leukaemia (ALL) is a fast-growing and life-threatening cancer that affects the blood and bone marrow, specifically impacting white blood cells (lymphocytes). Relapsed ALL comes back after treatment, and refractory ALL does not respond to initial treatment. Despite multiple available therapeutic options, this condition is associated with significant mortality and a poor survival rate.

Aucatzyl is a genetically modified autologous chimeric antigen receptor (CAR) T-cell therapy, a type of personalised cancer immunotherapy that is based on collecting and modifying the patient's own immune cells to treat their cancer. The modified T cells attach to and kill the cancer cells, thereby helping to clear the cancer from the body.

The recommendation is based on the results of a single arm, open-label trial (FELIX study) in 113 patients. About 64% of patients had a durable response (a period without disease signs or symptoms after treatment) with a median duration of 14 months. Around 49% showed a complete response, meaning the signs of cancer disappeared.

The most common observed side effects include cytokine release syndrome (a potentially life-threatening condition that can cause high fever, vomiting, shortness of breath, pain, and low blood pressure), immune effector cell-associated neurotoxicity syndrome (a condition that includes problems with use of language, seizures, headache, hallucinations, and mental confusion), and infections. Monitoring and mitigation strategies for these side effects are described in the product information and in the risk management plan.

In its overall assessment of the available data, the Committee for Advanced Therapies (CAT), EMA's expert committee for cell- and gene-based medicines, found that the benefits of Aucatzyl outweighed the possible risks in patients with ALL. The Committee for Medicinal Products for Human Use (CHMP), EMA's human medicines committee, agreed with the CAT's assessment and positive opinion, and recommended approval of this medicine.

Aucatzyl was supported through EMA's PRIority MEdicines (PRIME) scheme, which provides early and enhanced scientific and regulatory support to medicines that have a particular potential to address patients' unmet medical needs.

Aucatzyl is recommended for a conditional marketing authorisation. This type of approval allows the Agency to recommend a medicine for marketing authorisation with less complete data than normally expected, if the benefit of a medicine's immediate availability to patients outweighs the risk inherent in the fact that not all the data are yet available. In order to confirm the safety and efficacy of Aucatzyl, the company has been requested to submit long-term follow-up results of the FELIX study, and to conduct a non-interventional study based on a patient registry.

The opinion adopted by the CHMP is an intermediary step on Aucatzyl's path to patient access. The opinion will now be sent to the European Commission for the adoption of a decision on an EU-wide marketing authorisation.

Image: Freepik

New guideline on inclusion of pregnant and breastfeeding individuals in clinical trials

June 4, 2025

EMA has released for public consultation a new guideline[1] providing recommendations on how to include and/or retain pregnant and breastfeeding people in clinical trials. The goal is to ensure developers generate robust clinical data in those populations, so that these individuals and their healthcare providers can make informed, evidence-based decisions when using medicines.

This guideline, developed jointly by global regulators and medicines developers through the International Council for Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH), marks a change in paradigm in the development of medicines in pregnancy and breastfeeding. It highlights that in principle, including pregnant and breastfeeding people in clinical trials should be considered for all medicines intended for people who can potentially give birth to children. It lays out the principles and conditions that should be met to ensure the safety of clinical trial participants, as well as their foetuses and babies.

Currently, pregnant and breastfeeding people are often excluded from clinical trials and those who become pregnant while participating in a clinical trial are frequently discontinued from the clinical trial. Less than 0.4% of all clinical trials currently submitted in the EU include pregnant people, and this falls to 0,1% regarding lactating individuals, according to data from the Clinical Trials Information System (CTIS).

As a result, product leaflets usually lack details about the benefits and risks of a medicine specifically in pregnancy and breastfeeding, requiring patients and healthcare professionals to make treatment decisions without this essential information. This can lead to suboptimal treatment decisions and potential harm. Meanwhile, the vast majority of pregnant people take medications, for example because of chronic diseases, infections, or pregnancy complications. The situation is similar in breastfeeding populations.

The guideline outlines the scientific and regulatory principles, as well as ethical considerations, for the inclusion of pregnant and breastfeeding individuals in clinical trials, both pre- and post-authorisation. It encourages proactive planning and early consultation of medicine developers with regulatory authorities to ensure the safety and efficacy of treatments during pregnancy and breastfeeding.

The guideline was open for consultation until September 15, 2025.

Reference

1. ICH E21 Guideline on inclusion of pregnant and breastfeeding individuals in clinical trials – Scientific guideline. Available from: https://www.ema.europa.eu/en/ich-e21-guideline-inclusion-pregnant-breastfeeding-individuals-clinical-trials-scientific-guideline

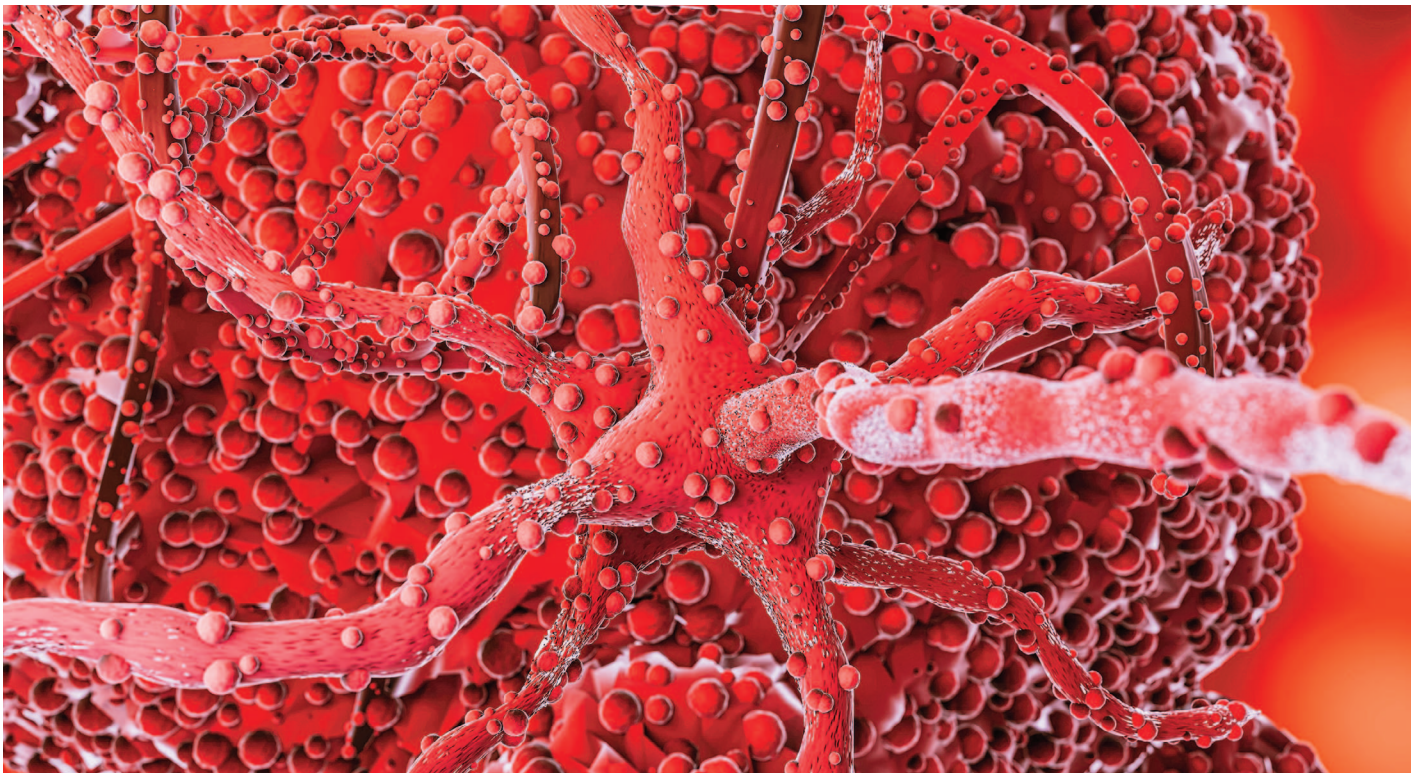


Image: Freepik

New stem cell therapy to treat patients with blood cancers

June 20, 2025

EMA has recommended granting a conditional marketing authorisation in the EU for Zemcelpro (dorocubicel/unexpanded umbilical cord cells) to treat adults with haematological malignancies (blood cell cancers). Zemcelpro can be used in patients requiring an allogeneic haematopoietic stem cell transplantation (allo-HSCT, transplantation of stems cell from a donor) following myeloablative conditioning (chemotherapy and/or radiotherapy) for whom no other type of suitable donor cells is available.

Haematological malignancies categorised depending on where they are first detected and include leukaemias (blood), lymphomas (lymph nodes), myelodysplastic syndrome and myelomas (bone marrow). They are frequently diagnosed cancers, and the only potential curative treatment option for several of these cancers is allo-HSCT. This type of transplant involves using donated stem cells to replace the recipient's bone marrow cells to form new bone marrow that produces healthy blood cells.

Stem cells used for transplantation are preferentially sourced from a matched donor, including a matched sibling or a matched unrelated donor. Umbilical cord blood cells can be used in patients who lack access to any type of suitable donor. However, the number of stem cells in umbilical cord blood is often low and can delay engraftment, the successful establishment and proliferation of the donor stem cells in the recipient's bone marrow.

Zemcelpro is a cell therapy containing stem cells from a donor's umbilical cord blood, some of which have been grown and multiplied (dorocubicel). By increasing the number of cells, Zemcelpro makes the stem cells from a small cord blood unit more effective.

The recommendation is largely based on a pooled analysis of two single arm, open-label studies which included 25 patients. In total, 21/25 (84%) patients achieved neutrophil engraftment (when donor stem cells successfully establish themselves in the recipient's bone marrow and produce neutrophils, a type of white blood cell) within a median time of 20 days, and 17 (68%) patients achieved platelet engraftment within a median time of 40 days.

The most common side effects observed in a wider pool of 116 patients treated with Zemcelpro include low levels of various types of blood cells and of antibodies that help fight infections, high blood pressure, infections, and engraftment syndrome, an inflammatory condition that can occur after HSCT. Acute graft-versus-host disease (GvHD), when donor/transplanted cells attack the body shortly after a transplant) up to 100 days after transplantation was reported in 60% of patients, and chronic GvHD appearing up to one year after transplantation was reported in 13% of patients. Monitoring and mitigation strategies for these side effects are described in the product information and in the risk management plan.

In its overall assessment of the available data, the CAT, EMA's expert committee for cell- and gene-based medicines, found that the benefits of Zemcelpro outweighed the possible risks in patients with haematological malignancies requiring allo-HSCT for whom no matched donor cells are available. The CHMP, EMA's human medicines committee, agreed with the CAT's assessment and positive opinion, and recommended approval of this medicine.

Zemcelpro was supported through EMA's PRIME scheme, which provides early and enhanced scientific and regulatory support to medicines that have a particular potential to address patients' unmet medical needs. Zemcelpro is recommended for a conditional marketing authorisation. In order to confirm the safety and efficacy of Zemcelpro, the company has been requested to submit long-term follow-up results of the single arm studies, conduct a randomised controlled study and a study based on a patient registry.

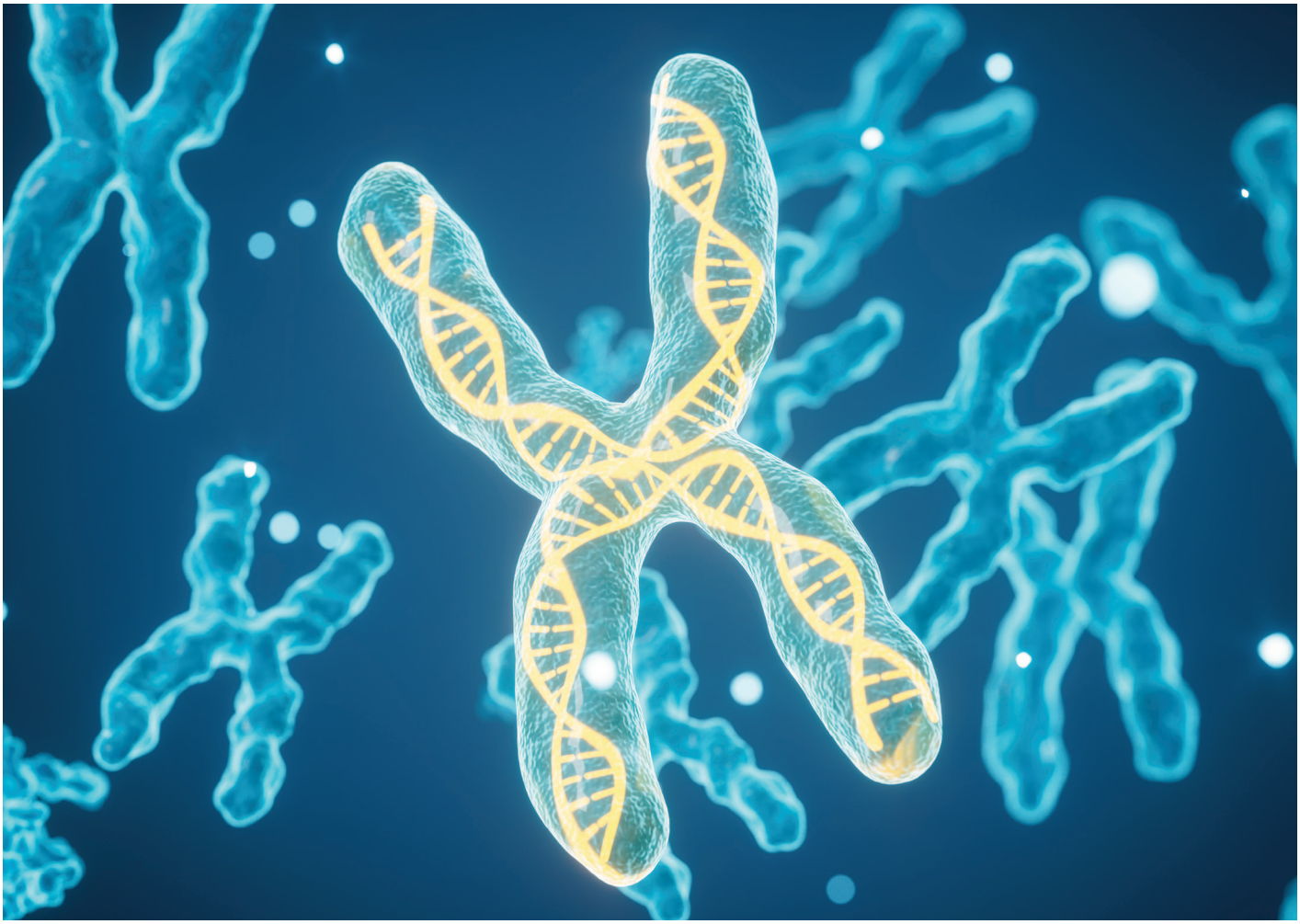


Image: Freepik

Strengthening supply chain of anti-D immunoglobulins

July 4, 2025

FMA and the Heads of Medicines Agencies (HMA), through the Executive Steering Group on Shortages and Safety of Medicinal Products (MSSG), have issued recommendations to address vulnerabilities in the supply chain of anti-D immunoglobulins.

These medicines are currently the only available treatment for the prevention of RhD immunisation during pregnancy. RhD immunisation happens when a pregnant person with RhD-negative blood type is exposed to RhD-positive blood from their foetus. This can lead to an immune reaction that can seriously impact the health of the foetus, and later of the newborn, and have potentially fatal outcomes.

Plasma, the liquid part of blood, collected from donors and containing the anti-D immunoglobulin is currently the only source for manufacturing these medicines. The numbers of donors are declining, and anti-D immunoglobulins are only produced in a limited number of countries, all located outside the EU. For this reason, the MSSG has been monitoring the supply chain of these medicines and has issued these recommendations to national regulators, the European Commission, as well as to the plasma industry and relevant research organisations, to support actions to strengthen their availability and prevent serious shortages.

EU Member States are recommended to create plans to secure the supply of anti-D immunoglobulins in the EU, guided by relevant safety, legal, ethical and regulatory aspects. These plans should also focus on reducing unnecessary use, for example through non-invasive pre-natal screening. Countries should support development and validation of alternatives to these medicines through research and funding and create prioritisation guidelines to manage shortages. In addition, they should implement communication campaigns to increase awareness of plasma collection for the development of plasma-derived medicinal products, such as anti-D immunoglobulins.

The European Commission is encouraged to identify measures to ensure supply continuity of these medicines and support and coordinate Member States' activities. Policy measures set out in the proposed Critical Medicines Act could be leveraged, such as joint procurement of manufacturing services to establish or increase supply of these medicines to the EU.

Finally, industry should ensure the adequate supply of anti-D immunoglobulins in Europe, including through investments in optimising manufacturing capacity and developing alternatives to plasma-derived anti-D immunoglobulins.

Anti-D immunoglobulins are included in the Union list of critical medicines; therefore, a stable supply of these medicines is considered vital for the functioning of EU health systems and the wellbeing of its citizens. While the recommendations address the anti-D immunoglobulin supply chain, the principles are also applicable to address vulnerabilities in the supply chain of other plasma-derived medicines.

Image: Freepik

New injection for easier prevention of HIV infection in the EU and worldwide

July 25, 2025

EMA has recommended granting a marketing authorisation in the EU for Yeytuo (lenacapavir) for pre-exposure prophylaxis (PrEP) in combination with safer sex practices to reduce the risk of sexually acquired human immunodeficiency virus type 1 (HIV-1) infection in adults and adolescents at high risk of becoming infected. PrEP is a cornerstone of HIV-control efforts in Europe and worldwide and is very effective at preventing infections if taken as prescribed. However, uptake and adherence are often suboptimal because access to some medicines is limited, and other available medicines require strict daily intake. Yeytuo will facilitate PrEP uptake and compliance because it only has to be administered twice a year via a subcutaneous injection. Of note, two tablets of Yeytuo on the first two days are required when starting the treatment, after which the medicine is given by injection every six months.

HIV-1 infection is of major public health significance. According to the WHO, in 2024 an estimated 1.3 million people became newly infected with HIV globally, including 160,000 new HIV infections in the European region and 650,000 in Africa, the region most affected by HIV.

HIV-1 impacts the body's immune system, particularly white blood cells that are important in helping to fight infections. If left untreated, HIV-1 infection can progress to acquired immune deficiency syndrome (AIDS), where the immune system is severely damaged, making the body vulnerable to opportunistic infections and some cancer types. Sexual intercourse is the most common mode of transmission of HIV-1.

Yeytuo contains lenacapavir, a first-in-class substance that binds to the proteins that make up the outer layer of HIV-1 (the capsid). By binding to these proteins, lenacapavir interferes with multiple steps in the HIV-1 lifecycle, thereby inhibiting viral replication, ultimately preventing HIV-1 infection.

CHMP's recommendation is based on the results of two randomised, double-blind, active-controlled, multinational trials. In the PURPOSE 1 trial, cisgender women, including pregnant and lactating women, between the age of 16 and 24 who have sex with cisgender males, were randomised in a 2:1 ratio to receive Yeytuo (n=2134) or Truvada (n=1068). At the time of the primary analysis, no new HIV-1 infections were observed in the Yeytuo group compared to 16 in the Truvada group.

In the PURPOSE 2 trial, men and gender-diverse persons from 16 years old who have sex with male partners, were randomised in a 2:1 ratio to receive Yeytuo (n=2179) or Truvada (n=1086). At the time of the primary analysis, two new HIV-1 infections were observed in the Yeytuo group compared to nine in the Truvada group. In both studies, participants who received Yeytuo showed higher adherence to their treatment than participants who received Truvada.

The most common side effects observed were injection site reactions, including pain and hard lumps (injection site nodules) that can persist for a long time or not disappear.

Yeytuo was evaluated by the CHMP, EMA's human medicine committee, under an accelerated timetable because it is considered to be of major public health interest in the EU and the rest of the world. The CHMP simultaneously reviewed the medicine for the EU market, under the centralised procedure, and for non-EU countries, under the 'EU-Medicines for all' (EU-M4all) programme in collaboration with the WHO and the target countries. The CHMP scientific opinion under the EU-M4all procedure supports global regulatory capacity building and contributes to the protection and promotion of public health beyond the EU.

Image: Freepik

First reformulation of an inhaled medicine with environmentally friendly gas propellant

Jul 25, 2025

EMA has recommended a change in the composition of Trixeo Aerosphere and its duplicate product Riltrava Aerosphere to replace the existing gas propellant with a low global warming potential (GWP) gas alternative. The new low GWP alternative propellant has a 1000-fold reduction in global warming potential and similar physical properties compared to the current propellant.

Trixeo Aerosphere and Riltrava Aerosphere are the first inhaled medicines in the EU that have a gas propellant with low GWP. They are used for maintenance treatment in a subset of adults with moderate-to-severe chronic obstructive pulmonary disease (COPD) and are administered as two inhalations twice daily using a metered dose inhaler (MDI).

A critical component of the formulation of an MDI is the propellant (liquified compressed gas) that generates an aerosol cloud containing the small particles of active pharmaceutical ingredients that are then inhaled by the patient.

High GWP gases, including hydrofluorocarbon gases such as the propellants used in pressurised MDIs treating respiratory diseases, are being phased out for environmental reasons in line with the current EU Regulation on fluorinated greenhouse gases (EU Regulation 2024/573), and applicable legislation in other regulatory constituencies. The marketing authorisation holder for Trixeo/Riltrava Aerosphere investigated replacement options for the current propellant, with a focus on a lower GWP propellant that could maintain the same performance properties for the medicinal product.

The reformulated Trixeo/Riltrava Aerosphere with the same active ingredients and dose has been characterised in line with the principles outlined in the draft Guideline on the requirements for demonstrating therapeutic equivalence between orally inhaled products for asthma and COPD and is therapeutically equivalent (i.e., works the same way and gives the same results in the lungs and the body) to the product currently on the market. Studies have confirmed that the safety and efficacy of the reformulated medicine are equivalent to those of the currently approved product.

The opinion will now be sent to the European Commission for the adoption of a decision on the variation to the marketing authorisation.

Gained in Translation

SECTION EDITOR
Ana Sofia Correia
ana@anasofiacorreia.com

Guidelines for quality translation of patient information materials from English to Russian

Ekaterina Chashnikova
Freelance medical writer and translator
Bar, Montenegro

doi: 10.56012/ejmo5136

Correspondence
Ekaterina Chashnikova
chashnikova.e@gmail.com

Abstract
Accurate and accessible translation of patient materials can have a considerable impact on health outcomes, especially in settings where the official country language differs from the patient's native language. Translations also play a crucial role in patient and caregiver education in the context of rare diseases, where materials in the native language may be limited. This article discusses existing requirements for patient-oriented texts and looks into the current plain language recommendations for Russian. The second part of the article features multiple excerpts from real-life English-to-Russian translations of patient-oriented materials and discusses suggested edits that facilitate the patient's understanding of the text.

The translation of patient information is highly relevant in the age of globalisation, accessibility, international clinical trials, and increased migration in many parts of the world. With Russian being the ninth most spoken language in the world, with approximately 255 million speakers across different countries,[1] high-quality medical translation is crucial for ensuring healthcare access for Russian-speaking patients. Moreover, accurate and accessible translations of patient materials can significantly impact health outcomes, and directly influence treatment adherence. This article discusses the translation of patient information materials from English into Russian, including the types of materials, text requirements, readability tools, and common translation errors. Since Russian uses the Cyrillic alphabet, the Russian expressions and sentences were Romanised according to ISO-9 standard to help the journal readers who do not speak Russian read the examples.

Types of patient information materials
Patient information materials include brochures and leaflets about diseases or treatment options; books about medical conditions for patients and caregivers; promotional materials about treatment methods or new medications; package inserts (patient information leaflets); informed consent forms and patient information sheets for clinical trials; information on websites and mobile apps for patients. This article describes the principles that apply to all kinds of patient-oriented materials and illustrates them with quotes from translations of books and brochures for patients.

Text requirements for patient materials
Many guidelines for writing information materials for patients are universal and can be applied across different languages. For example, the American Academy of Ambulatory Care Nursing published its *Guidelines for Developing Patient Education Materials*,[2] which include the following recommendations:

- Short sentences (no more than 10 words)
- Reading level appropriate for grades 3 to 5 (most guidelines recommend grades 5-8)
- Each paragraph containing no more than 2 to 3 sentences and expressing a single concept
- Bullet points and numbered lists for clarity
- Visual terms that patients can understand (e.g., "runny nose", "redness")
- Include approximate or exact measurements as illustrations (e.g., "pain present for more than 30 minutes")
- Plain language
- Definitions of complex terms
- Use of familiar words

The Russian requirements for patient-oriented medical educational materials echo these principles. According to a guideline published in 2007,[3] the writing style should be:

- Clear and concise
- Free of specialised medical terminology
- Adapted to the general educational level of patients

The reason these guidelines are universal is that they are based on general plain language principles.[4,5] Some of these principles are especially important for translation from English into Russian:

- **Directly addressing the reader**: There is a certain predisposition to use overly formal or bureaucratic language and impersonal sentences in Russian. Translators may also avoid addressing the reader with the formal "you" if they are used to translating specialist documents.
- **Using active voice**: Passive voice complicates text comprehension, but medical translators might prefer it if their prior experience was focused on technical and scientific texts.
- **Avoiding abbreviations and acronyms**: It is not typical for Russian texts to introduce acronyms as often as in English texts, but some abbreviations may be inadvertently carried over during translation.
- **Avoiding complex sentences**: The tendency for complex sentences in Russian can lead translators to combine short plain-language

English sentences into longer, more complicated Russian sentences. An example is provided later in this article.

- **Giving clear instructions**: For example, write "do this" instead of "one should do this". Again, the tendency to write impersonal sentences can outweigh the recommendation to provide direct, clear guidance.

Readability assessment

Standardised readability criteria help overcome the predisposition to long words, passive voice, and complex sentences. Translators can use Microsoft Word's readability statistics feature,[7] which is based on Flesch Reading Ease[6] and Flesch-Kincaid formula, or a free online tool at http://ru.readability.io/ based on plain language guidelines.[8]

Another free automated tool can be found at https://www.plainrussian.ru/. This tool provides readability assessments based on the Flesch-Kincaid formula, Coleman-Liau index, Automatic Readability Index, SMOG (Simple Measure of Gobbledygook), and Dale-Chall formula, all adapted for Russian. The website includes links to detailed descriptions of these indexes.

When using the automated readability tools, remember that medical texts for patients can still include medical terms and expressions that might trigger false alerts in such tools.

Common translation errors and solutions

This part of the article illustrates the aforementioned plain language principles and writing guidelines with examples from real translations. In most cases, only one or two edits are discussed to highlight a specific point. Readers may also notice additional edits to improve readability, but these are not discussed to avoid repetition.

Example 1: Age-appropriate language

When translating materials for a specific age group (such as teenagers), it is important to use appropriate language and forms of address. In Russian, using the informal address "ty" rather than the formal "vy" for teenagers creates a more relaxed tone when discussing sensitive topics. This approach is used in the Association of Clinical Trials Organisations templates for informed consent forms to be used in paediatric trials[9] and in Russian medical books for children.

The example above is from a brochure for teenagers with cystic fibrosis (CF). The English text is clear and light, while the original translation is less accessible. Using the informal form of address helps the reader feel at ease and engage with complex topics, such as puberty and growing up with CF.

Source	Original translation	Edited translation
As a young person with cystic fibrosis (CF) you might feel that you sometimes have double the troubles. CF can certainly complicate matters. Having CF means that staying well and staying in charge of your life can be hard work – hard work that needs to happen daily. Nobody can do it for you, and we all know it's not much fun.	Kak lyuboj chelovek s diagnozom mukoviscidoz, vy *ispytyvaete* gorazdo bolee ser'eznye slozhnosti. *Zabolevanie mukoviscidoz*, konechno, oslozhnyaet mnogie voprosy. Pri mukoviscidoze podderzhanie xoroshego samochuvstviya i nesenie otvetstvennosti za svoyu sobstvennuyu zhizn' uzhe schitaetsya tyazhelym povsednevnym trudom. Nikto ne smozhet sdelat' e'tu rabotu za vas, i vse my osoznaem, naskol'ko e'to slozhno.	Kak molodoj chelovek s diagnozom mukoviscidoz, ty ispytyvaesh' gorazdo bolee ser'eznye slozhnosti. *Mukoviscidoz*, konechno, mnogoe uslozhnyaet. Pri mukoviscidoze podderzhanie xoroshego samochuvstviya i upravlenie svoej zhizn'yu mogut stat' tyazhelym povsednevnym trudom. Nikto ne smozhet sdelat' e'tu rabotu za tebya, i vse my ponimaem, chto e'to neprosto.

Example 2: Sentence length

Here, the source text contained four short sentences, which the translator combined into two longer ones using conjunctions. The second sentence included two conjunctions and a colon. The edited version restores the original four sentences for easier comprehension.

Source	Original translation	Edited translation
Young people with reduced height and weight and/or delayed puberty can have a hard time. You may look younger than your friends. Worse, you may be treated as if you are younger than you actually are. Sometimes that's not easy to deal with.	Molodye lyudi s deficitom rosta i vesa i/ili zapozdalym polovym sozrevaniem mogut ispytyvat' bol'shie trudnosti v soc. obshhenii. *Obychno vy vyglyadite molozhe, chem vashi druz'ya, bolee togo, vas schitayut mladshe, chem vy est' na samom dele: inogda e'to vyzyvaet bol'shie slozhnosti.*	Molodye lyudi s deficitom rosta i vesa i/ili zaderzhkoj polovogo sozrevaniya mogut ispytyvat' bol'shie trudnosti. *Ty mozhesh' vyglyadet' mladshe svoix druzej. Xuzhe togo, s toboj mogut obrashhat'sya, kak s mladshim. Inogda s e'tim nelegko spravit'sya.*

Example 3: Common expressions

This translation illustrates the principle of using expressions that patients commonly use in everyday life. In Russian, the standard expression for "attending doctor" ("your CF doctor" in the text) is "lechashhij vrach", not "lechashhij doktor". By using familiar expressions, the translator helps the reader free up cognitive resources to focus on important concepts, such as the pathogenetic mechanisms of the disease or the mechanism of action of a medicinal product.

Source	Original translation	Edited translation
If your growth is delayed, ask *your CF doctor* what can be done to speed it up.	Esli vy zametili, chto process rosta zapazdyvaet, pointeresujtes' u svoego *lechashhego doktora*, kakim obrazom mozhno ego uskorit'.	Esli u tebya zaderzhka rosta, pointeresujsya u *lechashhego vracha*, kak mozhno ego uskorit'.

Example 4: Avoiding redundant pronouns

In Russian, possessive pronouns are often unnecessary and make the text sound redundant and unnatural.

Source	Original translation	Edited translation
If you leave it untreated, you might be reluctant to cough and clear *your lungs*, which may lead to more chest infections.	Esli ostavit' e'to bez lecheniya, vy mozhete nachat' rezhe kashlyat' i prochishhat' *svoi legkie*, chto mozhet vyzvat' rost chisla infekcij dyxatel'nyx putej.	Esli ostavit' e'to bez lecheniya, ty mozhesh' nachat' rezhe kashlyat' i prochishhat' *legkie*, chto uvelichit chislo infekcij dyxatel'nyx putej.

Example 5: Simplifying complex sentences

The edited version reduces the use of conjunctions and avoids repetitive wording, making the sentence easier to understand.

Source	Original translation	Edited translation
You need to know that if you are on antibiotics, there is a higher chance of failure of the pill, especially when you change antibiotics.	Vy dolzhny znat', *chto esli* vy prinimaete antibiotiki, to povyshaetsya veroyatnost' *togo, chto* tabletka ne podejstvuet, osobenno pri smene antibiotikov.	Ty dolzhen znat', *chto* pri prieme antibiotikov, osobenno pri ix smene, tabletki s bol'shej veroyatnost'yu ne podejstvuyut.

Example 6: Reflexive verbs

Reflexive verbs reduce readability in Russian and should often be replaced with the active voice using a subject or with impersonal constructions, as shown in the example above.

Source	Original translation	Edited translation
But to most people, the word chemotherapy means medications used for cancer treatment. It's *often shortened* to "chemo."	Odnako dlya bol'shinstva lyudej mir ximioterapii ogranichen preparatami, ispol'zuemymi dlya lecheniya raka. Chasto *ona sokrashhaetsya* prosto do «ximii».	Odnako dlya bol'shinstva lyudej slovo ximioterapiya oznachaet preparaty dlya lecheniya raka. Chasto *ego sokrashhayut* do «ximii».

Example 7: Punctuation for concise text

Here, using a hyphen instead of a subordinate clause simplifies sentence structure.

Source	Original translation	Edited translation
You and your cancer doctor, called an *oncologist*, will decide what medication or combination of medications you will get.	Vy i Vash lechashhij vrach, *kotoryj nazyvaetsya onkologom*, reshite, kakoj preparat ili kombinaciyu preparatov Vy budete poluchat'.	Ty i tvoj lechashhij vrach – *onkolog* – reshite, kakoj preparat ili kombinaciyu preparatov ti budesh poluchat'.

Example 8: Thinking outside the box

We can omit addressing the reader, as it is merely a figure of speech in this context. This makes the translation shorter and clearer.

Source	Original translation	Edited translation
Sometimes chemo is the only treatment you need. More often, chemo is used with surgery or radiation therapy or both. Here's why:	Inogda ximioterapiya mozhet stat' edinstvennym lecheniem, *kotoroe Vam potrebuetsya*. Gorazdo chashhe ximioterapiya naznachaetsya v sochetanii s xirurgicheskim lecheniem i/ili luchevoj terapiej, i vot pochemu:	*Nekotorym pacientam trebuetsya* tol'ko ximiya. No gorazdo chashhe ee provodyat v sochetanii s xirurgicheskim lecheniem i/ili luchevoj terapiej, i vot pochemu:

Example 9: Simplified medical terms

You can see that the translator did not shorten "chemotherapy" to "chemo" in Russian, even though a shortened version was introduced earlier and used consistently in the source text. Here are some other examples of simplified Russian terms:

- Arterial'noe davlenie → davlenie [English equivalent: arterial pressure → pressure]
- Vakcinaciya → privivka ["privivka" is a colloquial equivalent of "vaccine / vaccination"]
- Vakcinirovat' → sdelat' privivku [same here for the verb "to vaccinate"]
- Intraabdominal'no → v bryushnuyu polost' ["intraabdominal" replaced with expression that translates as "into abdominal cavity" and does not use a Latin-derived word]
- Peroral'no → vnutr' [again, replacing a Latin-derived equivalent of "per os" with a simpler synonym]
- Oftal'mologicheskij → glaznoj [ophthalmological → eye]
- ChSS → pul's [replacing a common abbreviation of "frequency of heart rate" with "pulse"]
- Uroven' glyukozy v krovi → saxar/uroven' saxara [glucose blood level → sugar / sugar level]

Note that the possibility of some of these simplifications depends on context. For example, you cannot automatically replace "heart rate" with "pulse", but it may be appropriate in many situations.

Source	Original translation	Edited translation
Doctors take these factors into account, along with information published in medical journals and textbooks describing the outcomes of similar patients treated with *chemo*.	Vrachi uchityvayut e'ti faktory naryadu s informaciej, opublikovannoj v medicinskix zhurnalax i uchebnikax, v kotoryx opisyvayutsya rezul'taty, poluchennye u takix zhe bol'nyx, primenyavshix ximioterapiyu.	Vrachi uchityvayut e'ti faktory naryadu s rezul'tatami *ximii* v poxozhix sluchayax, opublikovannyx v medicinskix zhurnalax i uchebnikax.

Example 10: Avoiding overly formal language

The edited version replaces the heavy phrase "e'to oznachaet, chto" with the simpler "poe'tomu", eliminating two conjunctions and making the text easier to understand. Additionally, the last part of the sentence was simplified by replacing a verbal noun with a strong, active verb.

Source	Original translation	Edited translation
However, chemo medications can't tell the difference between healthy cells and cancer cells. *This means* normal cells are damaged along with the cancer cells, *and this causes side effects.*	Odnako ximiopreparaty ne razlichayut zdorovye i rakovye kletki. *E'to oznachaet, chto* normal'nye kletki povrezhdayutsya vmeste s rakovymi kletkami, *chto obuslovlivaet razvitie pobochnyx e'ffektov.*	Odnako ximiopreparaty ne razlichayut zdorovye i rakovye kletki. *Poe'tomu* vmeste s rakovymi kletkami povrezhdayutsya normal'nye kletki, *i voznikayut pobochnye e'ffekty.*

Conclusion

When translating patient information materials from English into Russian, focus on following plain language guidelines, steering clear of legal-style formulations, and using simplified terms or everyday expressions instead of specialised medical terminology whenever possible. Use online tools to assess readability and refer to formal guidelines.

These principles help produce translations that are both accurate and accessible, making complex medical information easier for patients to understand. Remember that plain, easy-to-read, or simplified language – whatever you call it – frees up the reader's cognitive resources to learn new information and make informed decisions about their health, rather than struggle with long sentences and overly complex terminology.

Acknowledgements

The translation examples were taken from texts that Ekaterina edited for a charity organisation Future Actually. These examples were also featured by Ekaterina in a webinar for UTIC in 2018.

References

1. Statista. The most spoken languages worldwide in 2023 [cited 2025 Apr 04]. Available from: https://www.statista.com/statistics/266808/the-most-spoken-languages-worldwide/
2. American Academy of Ambulatory Care Nursing. Guidelines for developing patient education material [cited 2025 Apr 04]. Available from: http://www.aaacn.org/guidelines-developing-patient-education-materials
3. Preparation of information and methodological support for the system of medical and health education and healthy lifestyle training for the population of the Russian Federation (Moscow, 2007). [cited 2025 Apr 04]. Available from: https://pandia.ru/text/78/097/64361.php
4. ISO 24495-1:2023 Plain language. Part 1: Governing principles and guidelines [cited 2025 Apr 04]. Available from: https://www.iso.org/standard/78907.html
5. Plainlanguage.gov. Federal plain languages guidelines. March 2011 [cited 2025 Jul 21]. Available from: https://www.plainlanguage.gov/media/FederalPLGuidelines.pdf
6. Orfogrammka. Readability Index (Flesch Reading Ease) [cited 2025 Apr 04]. (In Russian.) Available from:https://orfogrammka.ru/шпаргалки/индекс_удобочитаемости_флеша/
7. Microsoft Word readability statistics, Russian version [cited 2025 Apr 04]. Available from: https://support.microsoft.com/ru-ru/office/получение-статистики-по-удобочитаемости-и-уровню-документа-85b4969e-e80a-4777-8dd3-f7fc3c8b3fd2
8. What is "Plain Russian Language" from a technology perspective? Let's look at text readability metrics [cited 2025 Apr 04]. Available from: https://habr.com/ru/companies/infoculture/articles/238875/
9. ACTO templates for Patient Information Sheets and Informed Consent Forms [cited 2025 Apr 04]. Available from: http://acto-russia.org/index.php?option=com_content&task=view&id=314

Author information

Ekaterina Chashnikova is a pharmacist and has worked as a medical translator since 2007 for both translation agencies and direct clients. Ekaterina writes a blog for medical translators and shares her expertise at online and offline events.

Medical Communications and Writing for Patients

Editorial

Dear All,

This edition of *Medical Writing* offers a summary of a really excellent Meet and Share hosted by the Communicating with the Public Special Interest Group (CwP SIG). This Meet and Share explored the legalities around the EU General Data Protection Regulation (GDPR), EMA Policy 0070, and the newly emerging AI legislation, all of which were beautifully explained by Veronica K. Contreras, who is an expert in data protection, cybersecurity, and AI.

Together, GDPR, EMA Policy 0070, and the evolving AI legislation aim to advance scientific research, protect individuals' rights, and promote public health by fostering a well-informed and responsible approach to data management and technology use. Medical writers play a crucial role in ensuring compliance with these laws and regulations.

I'm incredibly grateful to Veronica for sharing her experience and knowledge so thoughtfully, and for answering all of our questions with such grace, patience, and humour! This is certainly a rapidly evolving field, and it takes a lot of time and effort to keep up.

I hope that you enjoy Veronica's article as much as I did, and in the meantime, stay safe and sane – enjoy the sunshine (if you have any!), and see you in the December issue!

Bestest, **Lisa**

SECTION EDITOR

Lisa Chamberlain James
lisa@trilogywriting.com

Communicating with the Public Special Interest Group Meet and Share: An overview of the EU General Data Protection Regulation, the EMA Policy 0070, and how they relate to AI

Veronica K. Contreras
Veronica K. Contreras, P.C.

doi: 10.56012/olsq3875

Correspondence to:
Veronica K. Contreras
veronica@vkc-pc.com

On December 3, 2024, Veronica K. Contreras, P.C., a firm that specialises in data protection, cybersecurity, and AI consulting services, had President and Founder Veronica Contreras give a presentation to EMWA.

The focus of this presentation was to provide an overview on the European Union (EU) General Data Protection Regulation 2016/679 (GDPR), the EMA Policy 0070, Artificial Intelligence (AI), and practical considerations on how best to apply the various laws and regulations in day-to-day business activities.

GDPR overview

The GDPR applies to entities in the EU, and those outside the EU, offering goods or services to individuals who reside in the EU or monitor individuals' behaviour within the EU. Under the GDPR, it is important to understand key concepts, and their applicability, for complying with the regulation. Key concepts not only include core definitions under the regulation, but also account for key principles, and other compliance requirements, that companies need to consider when conducting business in the EU and using individuals' personal data as part of companies' business activities and operations, inclusive of conducting, and supporting, clinical research. Key definitions include:

- **Processing**, i.e., includes various activities such as data handling, data collection, data storage, use, and destruction of personal data;
- **Personal data**, i.e., information relating to an identified or identifiable person, including pseudonymised data (coded information such as a patient ID number);
- **Sensitive personal data**, i.e., special categories of personal data, such as biometric characteristics, genetic data, religious beliefs, racial origin, medical health, political opinions, and data of minors under 16;
- **Data controller**, i.e., an entity that determines how personal data are processed;
- **Data processor**, i.e., an entity that processes personal data as instructed by a data controller; and
- **Subprocessor**, i.e., an authorised third-party to carry out processing activities on behalf a data processor's behalf.

Key principles under GDPR are designed to protect individuals' personal data and limit how such data may be processed by companies. These principles include:

- **Lawfulness, fairness, and transparency**, i.e., personal data must be processed fairly, in ways that individuals would reasonably expect and based on a lawful basis;
- **Purpose limitation**, i.e., personal data must only be collected for a specific purpose and only what is necessary for that purpose;

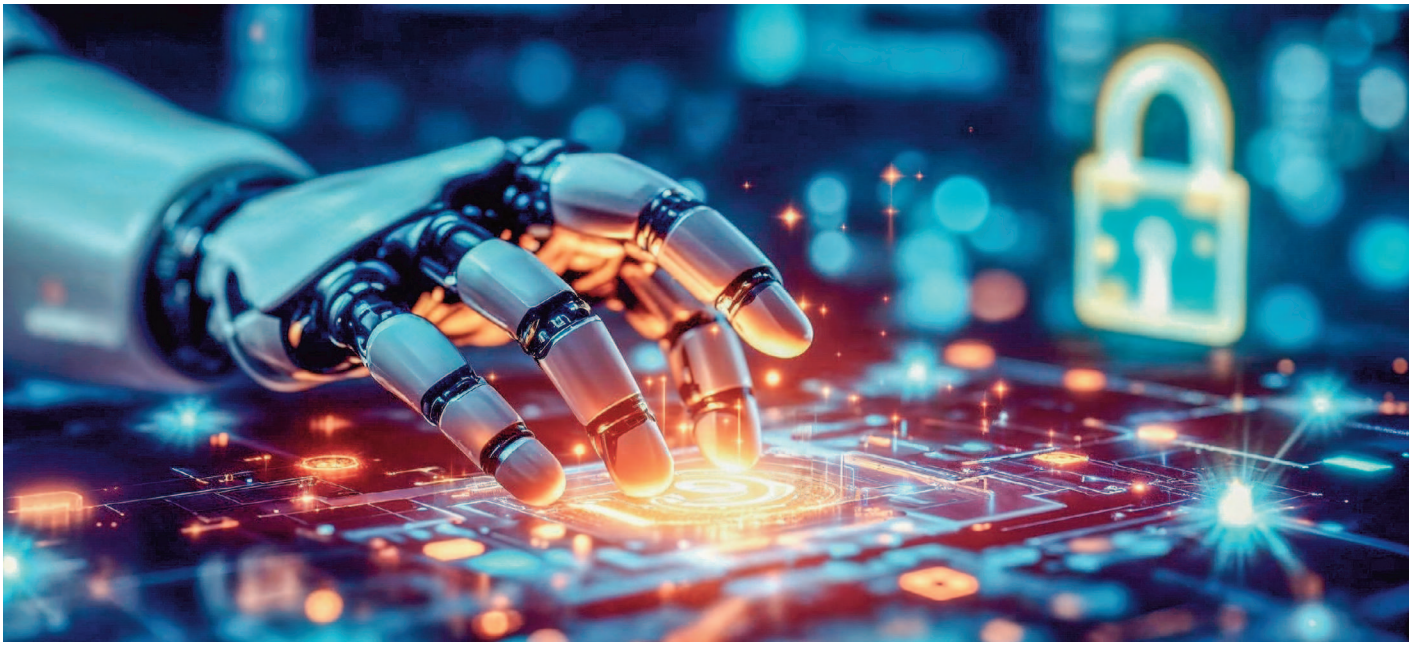


Photo: Freepik

- **Data minimisation**, i.e., ensuring that personal data collected are relevant, adequate, and limited to what is minimally necessary;
- **Accurate data**, i.e., personal data must be accurate, and necessary steps must be taken to update, rectify, or delete inaccurate data;
- **Data retention**, i.e., personal data must only be kept as long as necessary for the relevant processing activity; and
- **Data security**, i.e., implement appropriate security measures to protect personal data from unlawful or unauthorised processing, and from accidental loss, destruction, or damage.

The GDPR also incorporates requirements that any personal data processing must rely on a legal basis to allow for a processing activity to occur. These legal bases include:

- **Consent**, i.e., individuals must give clear and explicit consent to process their personal data for a specific purpose;
- **Contract**, i.e., processing is necessary for a contract with an individual or for human resource management activities;
- **Legal obligation**, i.e., processing is required to comply with legal or regulatory obligations;
- **Vital interest**, i.e., processing is necessary to protect an individual's life in emergencies;
- **Public interest**, i.e., processing is necessary for tasks in the public interest or official functions;
- **Legitimate interest**, i.e., processing is necessary for an entity's legitimate interests unless overridden by the need to protect personal data;
- **Archiving/scientific public interest**, i.e., processing supports archiving, scientific research, or statistical purposes;
- **Publicly available**, i.e., processing involves personal data intentionally made public by an individual; or
- **Permissible**, i.e., processing is otherwise allowed by applicable laws and regulations.

Some core compliance requirements under the GDPR afford individuals the ability to control how their data may be processed by companies and incorporates protective operational measures that need to be integrated into companies' operating practices. These core compliance requirements include:

- **Records of processing activities (ROPAs)**, i.e., data controllers and processors must maintain a ROPA log of all processing activities, document- ing contact details of the data protection officer, legal basis for the relevant processing activity, data categories, data recipients, international data transfers, data retention timelines, and data security controls;
- **Data processing agreements (DPAs)**, i.e., a DPA is required whenever a data controller uses a data processor, or a data processor uses a subprocessor, to process personal data. DPAs must include timing requirements for data breach reporting, data security controls, data transfer mechanisms, and indemnification and liability requirements;
- **International data transfer requirements**, i.e., personal data transfers to a third country must meet compliance requirements, including adequate data protection, data security controls, compliant data transfer mechanisms, and enforceable rights and legal remedies;
- **Individuals' rights**, i.e., the GDPR provides individuals with privacy rights, such as access to information, erasure, rectification, restriction of processing, data portability, objection to processing, and protection from automated decision-making and profiling; and
- **Data breach notification**, i.e., data controllers must notify relevant authorities and affected individuals within 72 hours of becoming aware of a data breach if it poses a high-risk to individuals.

EMA Policy 0070 was initially launched in 2015 to meet the growing demand for transparency in clinical data that forms the basis of regulatory decisions.

EMA Policy 0070 overview

The EMA Policy 0070 applies to pharmaceutical companies that have submitted clinical data as part of a marketing authorisation application or post-authorisation procedure for a human medicine in the EU. The policy enhances transparency and enables public access to clinical data, including clinical study reports (CSRs), clinical summaries, protocols, sample case report forms (CRFs), information on statistical methods used, and individual patient data (IPD).

Photo: Freepik

The EMA Policy 0070 was initially launched in 2015 to meet the growing demand for transparency in clinical data that forms the basis of regulatory decisions. This policy ensures that clinical data are published in an anonymised format to protect trial participants' identities and commercially confidential information.

The original policy had two phases: Phase 1 focused on CSRs, while Phase 2 was intended for IPD. The first publication was submitted in 2016. However, the policy was suspended in 2018 due to Brexit operational changes and revised in 2019 to cover both CSR and IPD. It resumed in 2020 with condensed reporting requirements for COVID-19 medicines. In 2023, the policy was relaunched, and as of September 2023, clinical data submitted for initial Marketing Authorisation Applications (MAAs) containing new substances with a Committee for Medicinal Products for Human Use (CHMP) opinion, were made public. Clinical data related to COVID-19, and other public health emergencies, continue to be made public. As of the policy relaunch in September 2023, the policy remains unchanged in content and has only undergone procedural changes. While step 2, of the policy relaunch, was originally anticipated in 2024, the EMA postponed next step requirements until 2025.

These policy background points highlight the evolution, and current status, of the policy, and emphasise its role in promoting transparency in clinical data. Some key requirements of this policy include:

- Submitting data in a format compatible with the EMA's publication system and within specified timeframes;
- All clinical data submitted for publication must be anonymised to protect patient privacy and commercially confidential information; and

Submission must comply with specific anonymisation guidance on how anonymisation should be carried out and the level of anonymisation required.

Key considerations to comply with the policy's anatomisation requirements and maintain patient privacy and confidentiality include:

- Understanding the process involved in transforming data into a form where individuals are no longer identifiable, and reverse engineering is impossible. If data are truly anonymised, they are no longer subject to data protection legislation requirements;
- Pseudonymisation reduces the linkability of a dataset with the original identity of an individual (e.g., patient ID). However, pseudonymisation alone does not result in an anonymous dataset, and data protection rules still apply. It is considered a best practice for enhancing security-related measures;
- Applicants/marketing authorisation holders are required to submit anonymous clinical reports. The EMA recommends a balanced approach to achieving adequate anonymisation, factoring in the risk of re-identification of a patient against the need to maintain data utility. For example, special consideration should be given to rare disease/small population studies by measuring the risk of re-identification and adapting anonymisation accordingly;

- Effective anonymisation considers three criteria:
  1. the possibility to single out an individual;
  2. the possibility to link records relating to an individual; and
  3. whether information can be inferred concerning an individual. If a planned report does not meet these criteria, an evaluation of associated re-identification risks must be performed; and
- Anonymisation techniques only extend to trial participants. Investigator, sponsor staff, and MAH applicant personal data should be redacted per EMA guidance.

AI legislation overview

AI legislation is on the rise, with new laws being passed to define legal requirements for AI use. These laws aim to protect individuals from fraud, theft, discrimination, bias, disinformation, and unintended consequences of AI use. Examples include the EU AI Act, and within the United States (U.S.), there are states, such as Colorado and California, that have passed their own AI legislation, which places significant obligations on developers, and providers, of high-risk AI systems (e.g., systems that make or significantly influence "consequential" decisions within the healthcare industry), including compliance with safety, transparency, fairness, algorithmic dis-

crimination prevention, and human intervention and accountability standards

AI system developers, and providers, must ensure robust evaluations are completed which address not only AI legislation requirements, but also consider data protection and information security requirements. These assessments should include, among other things, information about:

- An AI system that it will not affect individuals' safety and are thoroughly tested to ensure they are effective and not harmful to individual users;
- Algorithms used in AI systems will not discriminate against any individuals (e.g., gender, race);
- Any AI system use must be transparent, with clear documentation that includes descriptions of a system's features, general AI use, responsible parties, and explanations about AI outcomes;
- AI use should provide individuals the opportunity to opt-out from an AI system use in favour of a human alternative, where appropriate and applicable;
- What data will be used to train an AI model, inclusive of any personal data or other proprietary information;
- Cybersecurity measures enabled within an AI system; and
- Adherence to copyright laws.

Practical considerations

The aforementioned laws and regulations highlight the importance of transparency in clinical research and the interconnectedness of various laws in promoting public health, which is why it's important to understand how all these laws must be considered, and applied (where relevant), as part of any company's standard operational practices within the scientific and clinical research community.

For example, given that most clinical trials rely on patient consent for an individual to participate in a clinical study, GDPR requirements must be considered as part of the patient consent process. Under GDPR, consent often provides the legal basis which allows collection of a patient's personal data, inclusive of any medical records that would be needed as part of the applicable clinical study. Complying with good clinical practices and adhering to GDPR becomes a balancing act to lawfully process personal data, comply with data minimisation requirements, and avoid secondary use. A patient consent form, among other things, needs to include what data will be collected, why it will be collected, and how it will be collected. The challenge lies in ensuring that only the minimum data necessary are collected to meet the needs of a study and publishing goals, as outlined within the relevant consent form. GDPR limits how clinical data may be repurposed, or analysed, for future use, i.e., secondary use. Any future processing data uses that were not outlined in the relevant consent form will be prohibited unless patients are reconsented or the data are anonymised.

For any medical writers, or other stakeholders in the scientific community that share their research publicly, consideration should be given to what protections are in place to protect those individuals' intellectual property.

The EMA Policy 0070 complements GDPR by requiring all clinical reports to remove patient identifiers, thus anonymising all patient information and eliminating the ability to retrace an individual's identity. The process of anonymisation removes GDPR requirements because fully anonymised data are no longer considered personal data. This allows companies to not only comply with requirements under the EMA Policy 0070 (anonymisation requirements) but also leverage data for other use cases, such as data aggregation activities, without having to potentially re-consent patients to use their data**.** This is why compliance with the EMA Policy 0070 is valuable, advantageous, and promotes transparency and other benefits to advance public health.

While AI has been used to support scientific research for several years and more companies are integrating this technology into other aspects of clinical research to expedite and improve efficiencies when conducting clinical studies and publishing research for public use, it is important to understand how this technology may leverage a person's intellectual property to train an AI system. The basis of creating, or developing, an AI system requires certain information to train an AI model. Developers sometimes will look to public sources, such as clinicaltrials.org, or PubMed publications, to train their AI models. For any medical writers, or other stakeholders in the scientific community that share their research publicly, consideration should be given to what protections are in place to protect those individuals' intellectual property. Some considerations include:

1. whether research should be made commercially available beyond the EMA Policy 0070 requirements;
2. what protections a company like PubMed offers to protect individuals' intellectual property;
3. whether the "fair use" principle under copyright law is allowable or avoidable;
4. what royalty arrangements are available if an author's entire publication is used within an AI system;
5. require author acknowledgment labelling as part of an AI system; and
6. consider only sharing publication materials as part of an online subscription arrangement.

Conclusion

In the EU, GDPR sets the foundation for data privacy by defining key principles, and legal bases, for processing personal data, while the EMA Policy 0070 enhances transparency in clinical research by requiring anonymisation of clinical data. AI legislation is evolving to address the challenges, and risks, associated with AI use, emphasising safety, effectiveness, transparency, and algorithmic discrimination protections. Together, these laws and regulations, aim to advance scientific research, protect individuals' rights, and promote public health by fostering a well-informed and responsible approach to data management and technology use. Understanding and applying these interconnected laws in day-to-day responsibilities is crucial for compliance and achieving the overall objective of advancing public awareness and scientific research.

Author information

Veronica, president and owner of Veronica K. Contreras, P.C. (VKC-PC), specialises in data protection, cybersecurity, and AI consulting services, with more than 15 years of experience. VKC-PC assists clients with developing and implementing global data protection, cybersecurity, and AI compliance programmes, including without limitation: compliance with GDPR, Health Insurance and Portability and Accountability Act (HIPPA), California Consumer Privacy Act (CCPA), the Network and Information Security Directive (NIS2), and the EU AI Act.

Regulatory Matters

SECTION EDITORS

Clare Chang
clarechangphd@gmail.com

Zuo Yen Lee
zuoyen.lee@gmail.com

Editorial

Artificial intelligence is in the initial stages of adoption for regulatory medical writing, with the prospects of shifting the field from manual drafting to an era of intelligent automation and strategic content management. Here, Jenni Pickett and Vanessa de Langsdorff illustrate how AI-driven tools enable writers to transcend routine tasks – such as formatting and repetitive drafting – allowing them to focus on developing clear, compliant, and well-structured key messages for regulatory authorities. The authors emphasise that standardisation and modular content are now essential for achieving quality, consistency, and efficiency across global submissions.

Pickett and Langsdorff also highlight the evolving skill sets required: today's medical writers must have a foundation in AI technology, apply document content and data in the context of AI tools, and collaborate with both project teams and technology specialists. Successful adoption, they argue, requires not only technological fluency, but strong change management and alignment with organisational content standards. Far from replacing writers, AI elevates the profession, turning writers into content architects who guide document strategy and ensure scientific integrity.

Clare

The AI-enabled medical writer: A new era for regulatory writing

Jenni Pickett, Vanessa de Langsdorff
Yseop, New York, USA

doi: 10.56012/rjum8758

Correspondence to:
Jenni Pickett
jpickett@yseop.com

Abstract

As artificial intelligence (AI) becomes increasingly integrated into the pharmaceutical industry, regulatory medical writers find themselves at a critical intersection of science, language, and technology. The traditional approach to document authoring – manual, time-consuming, and highly variable – is being augmented by intelligent automation systems. This article explores how medical writers are navigating this transition, what technological concepts they must grasp to succeed, and how their roles are evolving to collaborate effectively with cross-functional tech teams. The future of regulatory writing is here, and it is structured, standardised, and AI-enabled.

Regulatory medical writers have historically expended significant time and effort to transform complex clinical data into clear, compliant narratives for health authorities. However, the introduction of AI tools into medical writing is rapidly altering their workflow. No longer confined to laborious manual document development processes, today's writers are expected to work in dynamic, tech-driven environments where content must be modular, reusable, and aligned with digital workflows.

The integration of AI into the writing process requires scale and standardisation to achieve real time savings:

- **Standardisation** limits individual preferences in writing style and formatting.
- Medical writers focus on **what to say**, the key message, and the significance of the data.
- Technology manages **how it is said**, ensuring consistency in style and structure.
- **Consistent content** improves quality control, streamlines the review process, and supports dossier assembly and regulatory approval.

> Writers have the opportunity to embrace technology and step into a more strategic role, becoming content architects who help standardise information from source data through to final review.

With global submissions, multiple indications, and mounting pressure to reduce time-to-market, medical writing teams are exploring automation tools that can handle content generation and reuse with less tedious intervention. This means learning not only how to create fit-for-purpose regulatory content, but also how to configure it for automated workflows – structuring content so it can be parsed, analysed, and reused across the full regulatory dossier. Writers have the opportunity to embrace technology and step into a more strategic role, becoming content architects who help standardise information from source data through to final review.

Bridging the gap between writing and technology

For writers to collaborate effectively with tech teams, they must understand the language and logic of the tools they're being asked to use. This begins with foundational knowledge of how AI

is implemented in regulatory writing tools.

Most medical writing software aligns with one or more of the following categories:

- Classic programming includes most software we have had up until 2019, like structured content authoring tools that store blocks of content to be reused between different documents.
- Symbolic AI, also called an expert system or deterministic AI, uses rule-based AI technology to generate accurate text and tables. Because symbolic AI requires an extensive knowledge base, it is typically found only in proprietary software designed for a specific task. Symbolic AI is able to perform Natural Language Generation (NLG) with 100% data accuracy.
- Generative AI creates text based on probabilities it has learned from huge datasets. Machine learning (ML) is used to identify patterns in human text, images, video, and audio. These patterns form a large language model, or LLM. Using a model to predict a response makes generative AI flexible to a variety of tasks, but its probabilistic nature means it can be challenging to get an exact reproducible result, and errors are a possibility.

Medical writing software may focus primarily on one underlying technology, or may be a blend of two or three. For example, the text could be generated by symbolic AI and then summarised or enhanced by generative AI. Classic programming provides you with buttons and menus to execute actions, for example.

Learning the tech lingo

To navigate AI tools confidently, medical writers must also become comfortable with some key technical terms:

- **Token**: The smallest unit of text processed by an LLM, often a word or part of a word. For example, according to the OpenAI Tokeniser, GPT-4o breaks "Learning the tech lingo" into 5 tokens: learning, the, tech, l, and ingo.
- **Context Window**: The total amount of text an LLM can "see" at once to generate accurate output. Regulatory documents can be too large to be processed by an LLM all at once because the context window may only be a few hundred pages.
- **Retrieval-Augmented Generation (RAG)**: A method that feeds relevant content into an LLM's prompt to improve factual accuracy. A RAG system breaks down long content into only the necessary chunks of information.
- **Multimodal Models**: AI systems that can interpret more than text – such as images, video, or audio. For regulatory writing, multimodal models are helpful to understand figures and schemas.
- **Hallucinations**: Content a large language model generates that is incorrect.
- **User acceptance testing**: A process where end users verify a system meets requirements before release. Testing typically includes predefined scenarios to validate accuracy, usability, and compliance.

Illustration: ChatGTP

This shared vocabulary enables smoother collaboration with product managers, developers, and data specialists – especially when evaluating or implementing software solutions.

Build versus buy: Choosing a software development strategy

Medical writers and their organisations face a critical decision in their AI adoption journey: whether to build a custom tool, buy a specialised platform, or adapt a general-purpose tool such as ChatGPT.

Illustration: ChatGTP

- **Custom tools (Build)** offer the highest level of control and integration but demand substantial investment, internal development resources, and ongoing maintenance for the life of the tool. The rapid evolution of AI technology adds further complexity to in-house development.
- **Specialised platforms (Buy)** are often designed specifically for medical writing and come with built-in security, compliance support, automation features, maintenance, and active user communities. These providers regularly enhance their tools, but adoption may require workflow changes and data-sharing agreements.
- **General-purpose AI tools (Adapt)** offer accessibility and flexibility making them appealing for experimentation. However, they pose challenges for regulatory writing, including data security risks, limited automation, lack of version control, and no safeguards against hallucinations. Prompting section-by-section may not yield major time savings, and lengthy regulatory documents often exceed these tools' context window.

Choosing the right approach depends on more than feature comparisons. It requires aligning with an organisation's goals, data security policies, and long-term scalability needs.

Working with tech vendors: what to expect

Successful implementation of AI writing tools depends on more than just choosing the right software. It also involves structured collaboration with technology vendors throughout evaluation, configuration, and deployment. The process typically begins with research and vendor outreach – understanding what's available, attending product demonstrations, and participating in hands-on evaluations.

Writers engaged in technology projects should expect to participate in pilot programmes, contribute to user acceptance testing, and provide real-world data and templates for evaluation. While sales and customer success teams usually lead these engagements, internal medical writing leads play a key role in reviewing content quality, assessing usability, and ensuring integration with existing workflows.

Participating in configuration and deployment also means adapting to software development processes. This might include working across different software environments – development, staging, and production – and providing structured feedback via tickets or feature requests.

Aligning content standards for AI readiness

One of the most important factors in successful AI adoption is the state of an organisation's content standards.[1] Many teams struggle with outdated templates, inconsistent formatting, or siloed writing styles. Without a shared approach to how key documents – such as Clinical Study Reports (CSRs) or summaries – present data and key messages, automation becomes significantly more difficult.

Content strategy

A clear, consistently implemented template supports automation and facilitates content reuse across the regulatory dossier. When content from study-level protocol and reports is structured for reuse, it can cascade into other documents such as summaries, investigator brochures, and briefing packages with minimal rework. This "intelligent content cascade" allows medical writers to shift their focus from redundant authoring to strategic messaging and scientific interpretation.

In large organisations, regulatory writing teams are often structured by function (e.g., clinical, safety) or therapeutic area, which can lead to a divergence in templates and inconsistent content standards. Readiness for automation requires realignment to common templates, style guides, and content standards.

An example of document readiness is the ICH M11 Clinical Electronic Structured Harmonised Protocol (CeSHARP) guideline and its accompanying Technical Specification document and Template.[2-4] These documents provide not only a common structure for all study protocols, but also a standard for electronic exchange of protocol metadata. At a minimum, consistent and descriptive document headings allow AI tools to easily find relevant content for intelligent reuse.

Data readiness

Typical clinical study output tables are designed to be human readable, with visual cues like merged column headers and indentations to indicate relationships that machines have difficulty understanding. Many AI tools work better with machine readable formats where the relationships of each data point to its descriptors or metadata are clearer.

A technology team may have questions about what format the data files come in, for example, SAS files, RTF tables in Word, or CSV tables in Excel. There are also other formats that structure data like JSON, HTML, and XML. Some medical writing AI tools can work with the raw individual-level Clinical Data Interchange Standards Consortium (CDISC) data that sponsors are already required to submit to health authorities. For example, individual patient data in Study Data Tabulation Model (SDTM) or Analysis Data Model (ADaM) formats can be used to write patient narratives.

In 2024, CDISC released a new data standard for analysed data called the Analysis Results Standard, in an effort to make the final analysed data, the summary statistics and endpoint analyses for example, more standardised and machine readable.[5] The push toward data standardisation is supporting this type of TLF output standard across sponsors.

At a minimum, a consistent format per type of study table (e.g., Overview of Adverse Events) across all studies allows for an AI tool to consistently and easily process tables into text.

Lean authoring: evolving to meet new demands

The advent of generative AI has wide implications on medical writing content standards. Lean authoring was originally developed with manual writing and human reading in mind. It was optimised at the individual document level, with an emphasis in saving writing, quality control, and reviewing time. Repetition of data from tables in the body text was intentionally minimised, as human readers could easily interpret patterns in tabular format, and creating and checking numbers in-text was resource-intensive to write and verify.

As the regulatory writing landscape shifts toward automation and global dossier strategies, lean authoring is adapting to meet new demands. Now, authoring long text with accurate data points is possible in seconds. However, enabling AI to write in a tightly controlled, concise format requires additional development effort.

To support an intelligent content cascade across the dossier, the goal shifts from minimalism to fit-for-purpose: the text must remain concise but also contain enough context to be understandable on its own. This is especially important as health authorities begin using AI systems to assist in their review. While human reviewers can easily interpret data in tables, large language models may struggle to determine which parts of a table are most relevant, how schemas should be interpreted, or how to follow contextual links. Including key data points and core messages directly in the text may improve the likelihood that AI systems summarise the information accurately.

Change management: preparing for a new role

It is tempting to apply new technology to existing ways of working, but embracing new approaches can accelerate the path toward AI-assisted submissions. A few practical steps can lay the groundwork for successful adoption:

- Revise document templates to guide standardisation, consistency, and fit-for-purpose lean writing
- Train teams on the benefits of automated text generation and standardised templates
- Update document preparation workflows, such as reviewing, locking, and quality checking data-independent content prior to database lock

For high-achieving professionals like medical writers, significant changes to long-standing work practices can be unsettling. Fears of becoming less valuable, of not being able to achieve career goals, or losing professional standing can lead to scepticism or resistance to new technology initiatives.

Successful technology initiatives prioritise change management to help dispel myths, empower users, and ease the workload strain during an impactful change. Thoughtful change management involves clear communication of how roles will change, affirmation of each contributor's value, small steps to meaningful and achievable goals, and a defined process for everyone to follow to success.

The role of the regulatory medical writer has evolved to suit the needs of the documents dramatically over time. We have progressed from circulating paper drafts to leading collaborative authoring and enforcing compliance with electronic Common Technical Document standards. The AI technology progression will guide us into the next evolution, that of becoming a configuration lead and editor.

Here is an example of how AI-enabled regulatory medical writers are already generating documents in a fraction of the time:

- Ahead of database lock, the writer leads the creation of a first draft as follows:
  - Gathers all the source documents, dry run data outputs (or shells), and the appropriate Word template and connects them to the draft
  - If required, transforms the data into a machine-readable format by ensuring each data point is linked to key terms required for text generation (e.g, if the mean age in the placebo group is 65 years, the data point 65 is linked to years, age, mean age, and placebo)
  - Reviews the AI configuration template to confirm that the appropriate data and sources are linked to each section and that the content plan aligns with the document purpose, making adjustments as required
  - Populates the data-independent sections (e.g, introduction, study design) with a single click, then uses options within the tool to add additional context or enhance the text as needed
  - Reviews the data-independent sections with the team and locks the sections after review is complete
    - Quality control can be performed for data-independent sections using functions within the tool that allow the reviewer to view the source text and any changes that were made
  - Reviews the data-to-text plan with the team and adjusts the plan as needed (e.g., describe, compare groups, create an in-text table)
- After final data is available, the writer completes the document with these last few steps:
  - If required, converts final data into a machine-readable format
  - Populates most data sections (e.g., disposition, safety) with a single click, then uses options within the tool to enhance the text as needed

- Matches efficacy tables to predefined endpoints and generates interpretive and descriptive text using the SAP as context
- Reviews the data sections with the team and locks the sections after review is complete
  - Quality control can be performed for data sections using functions within the tool that allow the reviewer to view the source table and see if any edits were made to the AI-generated text

Future-proofing your career

AI is often associated with efficiency gains, cost reductions, and concerns about job displacement. However, within the pharmaceutical sector, AI's primary value lies in augmenting capabilities, industrialising complex domain knowledge, and accelerating drug delivery to patients.

This context fosters a growing demand for professionals with expertise in both life sciences and digital technologies like AI and data science, leading to new roles such as:

- Creating and implementing AI tools
- Medical writer AI leads or super-users, who integrate AI in their daily workflows and train other users

In software development, medical writing expertise ensures AI tools align with regulatory requirements and medical writer needs. For instance, validating an adverse event analysis prompt requires medical and scientific knowledge. Medical writers now have opportunities as product owners, product managers, prompt engineers, and business analysts, leveraging their expertise to work in AI implementation teams.

For AI users, the writer's role evolves from crafting every line of text to configuring AI settings, reviewing machine output, and aligning AI with team and project needs. Automation increases the expertise required to evaluate and control AI-generated content. This seemingly paradoxical shift elevates the seniority of medical writing roles, with junior writers utilising AI for repetitive tasks.

Valued skills now include data interpretation, prompt engineering, and structuring document creation. Modular thinking and cross-functional collaboration are essential for regulatory and medical writers.

On the other hand, manual formatting, repetitive drafting, and versioning tasks are becoming less critical, as generative AI and automation tools assume these functions. The core value shifts to shaping messages, interpreting data, and ensuring quality.

Pharmaceutical companies are also embracing blended teams or "squads", which pair domain experts with technology specialists. These collaborative models enhance adoption of AI tools and ensure solutions are grounded in regulatory reality. Medical writers who develop fluency in AI tools and understand their strengths and limitations are well-positioned to become AI-enabled subject matter experts in their field.

Conclusion

The role of the regulatory medical writer is evolving. Writers are no longer just authors; they are collaborators in software development, stewards of quality content, and architects of intelligent content ecosystems.

By embracing foundational tech knowledge and advocating for clarity and standardisation, medical writers can lead their organisations through a successful digital transformation.

The future of regulatory writing elevates the value of medical writers from "hands on the keyboard" to that of a strategic content designer. Writers who adapt to AI not only preserve their relevance but expand their influence across the regulatory lifecycle and future-proof their career.

Acknowledgements

The authors would like to thank Elizabeth Curtin and Emmanuel Walckenaer for their review and ideas.

Disclosures and conflicts of interest

The authors are employed by Yseop, which develops AI tools for regulatory medical writing.

References

1. Nielsen IØ, de Langsdorff V, April J. How medical writing and regulatory affairs professionals can embrace and deploy generative AI at scale. Appl Clin Trials [Internet]. 2025 Jan 7 [cited 2025 Jun 11]. Available from: https://www.appliedclinicaltrialsonline.com/view/medical-writing-regulatory-affairs-professionals-embrace-deploy-generative-ai
2. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use. Clinical electronic structured harmonised protocol (CeSHarP): guideline M11 [Internet]. Draft version. Geneva: ICH; 2022 Sep 27 [cited 2025 Jun 11]. Available from: https://database.ich.org/sites/default/files/ICH_M11_draft_Guideline_Step2_2022_0904.pdf
3. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use. Clinical electronic structured harmonised protocol (CeSHarP): technical specification M11 [Internet]. Updated Step 2 draft. Geneva: ICH; 2025 Mar 14 [cited 2025 Jun 11]. Available from: https://database.ich.org/sites/default/files/ICH_M11_Template_Updated%20Step%202_ForReferenceOnly_2025_0203.pdf
4. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use. Clinical electronic structured harmonised protocol (CeSHarP): template M11 [Internet]. Updated Step 2 draft. Geneva: ICH; 2025 Mar 13 [cited 2025 Jun 11]. Available from: https://database.ich.org/sites/default/files/ICH_M11_Template_Updated%20Step%202_ForReferenceOnly_2025_0203.pdf
5. Clinical Data Interchange Standards Consortium. Analysis Results Standard (ARS) v1.0 [Internet]. Austin (TX): CDISC; 2024 Apr 19 [cited 2025 Jun 11]. Available from: https://www.cdisc.org/standards/foundational/analysis-results-standard/analysis-results-standard-v1-0

Author information

Jenni Pickett, PhD, serves as Medical Writing Customer Success Director at Yseop and has 16+ years of pharmaceutical experience. She is passionate about educating others about how to leverage AI and has presented multiple educational conference sessions to help medical writers learn principles, risks, and capabilities of AI.

Vanessa de Langsdorff is the Vice President, Customer Success at Yseop, where she leads delivery and customer success. After graduating from ESSEC Business School (MBA), and 15 years of experience in strategic and organisation consulting, she joined Yseop in 2018. She brings extensive experience in driving digital transformation for pharma companies leveraging AI.

Medical Devices

SECTION EDITOR

Payal Bhatia
payal.bhatia@icloud.com

Real-world data coming into play in the medical device world: What medical writers need to know

Crispin Bennett, Heidi Chapman, Laura C. Collada Ali
Trilogy Writing & Consulting

doi: 10.56012/rjum8758

Correspondence to:
Heidi Chapman
heidi.chapman@trilogywriting.com

Photo: Freepik

The integration of real-world data (RWD) into healthcare decision-making has transformed various aspects of clinical research, regulatory approval, and post-market surveillance. As the healthcare landscape continues to evolve, the use of RWD within the context of medical devices has garnered increasing attention from regulatory agencies, industry stakeholders, and the scientific community. With RWD now more available than ever before, clinicians and regulators alike are realising the benefits it offers: enhancing the relevance and applicability of research findings, improving patient outcomes and care, and supporting regulatory decision-making.

Medical devices, ranging from in-vitro diagnostics to therapeutic technologies, are subject to rigorous clinical assessments before reaching the market. However, while clinical investigations are essential for establishing initial safety and efficacy, these may not fully capture the complexity of real-world usage and long-term outcomes and may not be feasible for lower-risk devices. In Europe, the Medical Devices Regulation 2017/745[1] and In-Vitro Diagnostics Regulation 2017/746[2] introduce enhanced requirements for Post-Market Clinical Follow-up (PMCF) of all medical devices, requiring that data be continually collected and appraised throughout the entire lifetime of the device. In the United States, there are similar regulations and statutes (e.g., 21 CFR Parts 814 and 822 and the Federal Food, Drug, and Cosmetic Act Section 522) establishing post-market studies and surveillance requirements. In addition, in recent years, several key initiatives and publications have focused on harnessing RWD to complement traditional clinical trials and provide more comprehensive insights into the performance, safety, and effectiveness of medical devices in diverse, real-world settings. By leveraging RWD, researchers and regulators are better equipped to address gaps in evidence, improve post-market surveillance, and guide product development in a more patient-centred manner.

This article explores 3 publications and 1 initiative surrounding the use of RWD in the medical device sector that medical writers would benefit from being aware of. We identified these through a targeted literature search (see Table 1). These publications highlight key regulatory frameworks, emerging methodologies for data integration, and real-world case studies that demonstrate the transformative potential of RWD. Medical device writers need to be aware of these initiatives and proposals in order to

Real-world data (RWD)
refers to observational data collected from sources outside of traditional clinical trials, such as electronic health records, claims data, and patient-generated data used to understand patient health and healthcare delivery[3]. RWD offers a richer, more diverse data pool compared to controlled clinical environments.[3]

Real-world evidence (RWE)
is clinical evidence regarding the usage, and potential benefits or risks, of a medical product derived from the analysis of RWD.[3]

Table 1. Key messages of the analysed publications

Publication	Key messages	
McDermott O, Kearney B. The value of using real-world evidence as a source of clinical evidence in the European medical device regulations: a mixed methods study. Expert Rev Med Devices. 2024 Jan-Feb;21(1-2):149-163. doi:10.1080/17434440.2023.2291454 Epub 2024 Feb 4. PMID: 38041629.	• RWD for PMCF studies and clinical evaluation reports can aid in adhering to the European Medical Device Regulations. • Manufacturers can bridge gaps in clinical evidence by using RWD.	• RWE supplements clinical evidence from pre- and post-market clinical investigations, reducing the costs associated with these studies and supporting the manufacturer's benefit/risk conclusion. • How the medical device industry could utilise RWE and proposes an initiative in the EU similar to the FDA-sponsored NESTcc partnership.
Wang C, Rosner GL, Bao T, et al. Leveraging real-world evidence for determining performance goals for medical device studies. Stat Med. 2021 Dec 20;40(29):6577-6589. doi: 10.1002/sim.9199. Epub 2021 Sep 24. PMID: 34561895.	• The authors of this publication propose a methodology for integrating unstructured data (e.g., text, images) into regression models, addressing challenges like measurement error and high dimensionality. • They introduce a one-step estimation strategy that combines information retrieval and topic modelling to generate variables from unstructured data, which are then used in regression analysis.	• The one-step strategy substantially reduces bias in simulations. The method has quantitatively important effects in a leading application using Chief Executive Officer time-use data. The approach can be readily adapted by applied researchers. • Implications: This methodology enhances the ability to incorporate unstructured data into regression models, improving the accuracy and reliability of statistical inferences in various fields.
Shi L, Xuan D, Jakovljevic M. A review on the evolving environment of medical device real-world evidence regulation on market access in the USA. Cost Eff Resour Alloc. 2024 Oct 25;22(1):75. doi:10.1186/s12962-024-00582-9. PMID: 39456032; PMCID: PMC11515808.	• The FDA's CDRH oversees medical device approvals, which have evolved from the Medical Device Amendments (1976) to more recent policies, such as the 21st Century Cures Act (2016) and the Food and Drug Omnibus Reform Act (2022). • Since 2017, the FDA has increasingly recognised RWE/RWD as part of the evidence package for medical device approval, particularly for post-market surveillance, breakthrough devices, and alternative approval pathways. • Issues such as data availability, reliability, harmonisation, and interoperability remain barriers. The NEST was a database established to synthesise RWE from clinical registries, EHRs, and claims data.	• The future of RWE in medical device regulation will focus on improving data transparency, standardisation, and analytical methods to enhance regulatory confidence. Advancements in AI and machine learning will help analyse large-scale RWD, bridging the gap between clinical research and real-world applications. • Expanding RWE beyond safety and efficacy to include cost-effectiveness and patient-centred outcomes will further support its role in regulatory decision-making and health technology assessments. • While regulatory acceptance of RWE is growing, the disparities in affordability of the different treatments, connected with the related reimbursement policies may compromise patient access to innovative medical devices.

produce robust quality documents for regulatory submissions and post-market surveillance.

Real-world data and clinical evidence

A review has been published highlighting the opportunities that exist in medical device regulations for manufacturers of legacy devices to conduct real-world evidence (RWE) studies to bridge gaps in clinical evidence.[4] The primary value of RWE lies in its ability to provide an accurate and, therefore, more reliable measure of device safety and performance. This is due to the fact that often RWD mirror routine cases and provide bigger amounts of data compared to clinical studies run for market approval of a given device. RWE supplements clinical evidence generated from pre- and post-market clinical investigations, reducing the costs associated with these studies and supporting the manufacturer's benefit/risk conclusion.[4]

Relevance for medical writers

The regulatory framework of medical devices is continuously evolving. Medical writers need to be proactive in adapting to the changing frameworks that may leverage RWE more extensively in the future, with the possibility of having a clone initiative to the FDA-sponsored National Evaluation System for Health Technology (NEST – https://nestcc.org) partnership in the U.S. in other regions in the world. Medical writers should understand how RWE can

Table 1. cont.

Publication	Key messages
Baumfeld Andre E, Gee M, Magnus C, Greeman S, White P, de Mars M, Spring B. The Open Hand Initiative: Facilitating the use of real-world evidence in regulatory submissions through collaboration and transparency. Clin Pharmacol Ther. 2025 Apr;117(4):1072-1077. doi:10.1002/cpt.3539. Epub 2024 Dec 24. PMID: 39716998; PMCID: PMC11924146.	• RWD is increasingly used to support regulatory decisions, but its integration in IVD approvals remains complex due to data quality, availability, and regulatory clarity issues. The Open Hand Initiative was piloted to address these challenges in transitioning SARS-CoV-2 serology tests from EUA to full market approval. • Participants identified difficulties related to data access, standardisation, and inter-operability, as well as limitations in using retrospective data for regulatory submissions. The COVID-19 pandemic also introduced unique barriers, such as shifting diagnostic criteria, evolving patient populations, and rapid technological advancements in test development. • The initiative emphasised the need for clear regulatory expectations for RWE, improved data collection and validation processes, and better alignment between study design and FDA requirements. Ensuring data privacy compliance and leveraging existing RWD sources were also highlighted as critical factors for successful submissions. • The Open Hand Initiative demonstrated the potential for collaborative regulatory science, where industry and regulators work together to refine RWE standards. While initially focused on IVDs, the model could be expanded to broader medical devices and pharmaceuticals, improving the integration of RWD in premarket and post-market decision-making.

Abbreviations: CDRH, Center for Devices and Radiological Health; COVID-19, Coronavirus disease 2019; HER, electronic health record; EU, European Union; EUA, Emergency Use Authorization; FDA, Food and Drug Administration; IVD, In-vitro diagnostics; NEST, National Evaluation System for health Technology; RWD, real-world-data; RWE, real-world evidence.

Note: These 4 publications were identified as most relevant for medical writers involved in writing regulatory documents for medical devices upon the following search string on PubMed on Feb 24, 2025: ((“real-world data”[Title] OR “real-world evidence”[Title]) AND (“medical device”[Title] OR “in-vitro diagnostic”[Title])) Filters: from 2020 – 2025

((“real-world data”[Title] OR “real-world evidence”[Title]) AND (“medical device”[Title] OR “in-vitro diagnostic”[Title])) AND (2020:2025[pdat])

support the clinical evidence required under the Medical Devices Regulation, reduce costs, and contribute to ongoing benefit/risk profile evaluations, particularly, for legacy devices. They should also be aware that there are multiple sources of RWE within the medical devices world: from hospital charts to social-media listening, for example.

Using real-world data to define clinical investigations endpoints

A new method has been developed leveraging RWE to propose meaningful endpoints for assessing devices.[5] The method applies entropy balancing (a data processing method for matching treatment and control observations) to address possible patient dissimilarities between a study’s target patient population and existing real-world patients, taking into account operational differences between clinical studies and real-world clinical practice. Applying this method reduces the risk of biased conclusions to be drawn on a set of different sources, as it is often the case with RWD. The publication is technical but presents a practical case.

Relevance for medical writers

Even if medical writers are not responsible for the statistical methodology, it is important that they understand how RWE can be applied to aid selection of relevant performance and safety endpoints for device studies and assessments. Safety and performance parameters are used in different steps of the clinical evaluation of a medical device; e.g., when defining the state-of-the-art and when designing endpoints for a clinical investigation. However, it is challenging to determine performance or safety parameters when the amount of reliable, relevant, and available data are limited.

Medical writers need to be careful when using unstructured data to draw conclusions on the safety and performance data of a given device, as the risk for a biased conclusion is high. Weighting different formats and sources of data is not an easy task, but statistical methodologies applying new methods like the one presented in this publication can play an important role in producing a robust body of data, from which conclusions can be drawn safely. We should be aware of these methods and work in collaboration with statisticians to analyse and assess data. Adopting robust methodologies can ensure more accurate clinical evaluations and support the validity of regulatory documentation.

Medical writers need to be careful when using unstructured data to draw conclusions on the safety and performance data of a given device, as the risk for a biased conclusion is high.

Can RWD facilitate market access?

A recent review explores the role of RWE and RWD in the regulatory approval of medical devices within the U.S. FDA framework.[6] The review highlights key legislative milestones, challenges in integrating RWE into regulatory decisions, and potential future directions.

Relevance for medical writers

It is important for medical writers to be aware of the value of integrating RWE into submission dossiers for devices and what the regulatory framework is for doing this. This article provides valuable information on evolving guidelines and best practices for utilising RWE. By applying these medical writers can support teams in producing applications that draw more robust conclusions about the need for the device at hand.

An all-hands initiative: When manufacturers and regulators come together to leverage real-world data

The Open Hand Initiative presents an innovative approach to integrating RWE and RWD into the regulatory approval process for in-vitro diagnostic (IVD) devices.[7] The Open Hand Initiative, a collaboration between the FDA, device manufacturers, and the Medical Device Innovation Consortium (MDIC), promotes transparency by encouraging manufacturers to share insights from regulatory interactions to improve the quality and applicability of RWE.[7]

Relevance for medical writers

This article provides further valuable insights into the evolving regulatory landscape for medical devices, particularly the role of RWE in FDA submissions. Medical writers involved in regulatory documentation, clinical investigation design, and evidence synthesis can benefit from understanding how to align RWD collection with regulatory expectations to improve the quality of data and transparency of the types of data supporting submissions. Addressing regulators expectations in terms of the content of the documentation to be submitted may shorten time to response from the authorities and lead to a more successful submission.

Conclusion

As the regulatory landscape continues to shift, understanding the role of RWD in medical device development will be crucial for ensuring better patient outcomes and driving the future of healthcare technology. RWE use in the medical device sector continues to expand, and it offers valuable opportunities to strengthen clinical evidence, enhance post-market surveillance, and support regulatory submissions. Medical writers must stay informed of emerging methodologies, evolving regulatory frameworks, and collaborative initiatives like the Open Hand Initiative.[7] By doing so, they can help ensure that medical device documentation is robust, relevant, and aligned with current standards – ultimately contributing to more effective, evidence-based healthcare.

Acknowledgements

The authors would like to thank Julia Forjanic Klapproth for her valuable review of this text.

Disclaimers

The opinions expressed in this article are the authors' own and not necessarily shared by their employer or EMWA.

Disclosures and conflicts of interest

The authors declare no conflicts of interest.

References

1. Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC. Available from: https://eur-lex.europa.eu/eli/reg/2017/745/oj/eng
2. European Parliament and Council. Regulation (EU) 2017/746 of the European Parliament and of the Council of 5 April 2017 on in vitro diagnostic medical devices and repealing Directive 98/79/EC and Commission Decision 2010/227/EU. Available from: https://eur-lex.europa.eu/eli/reg/2017/746/oj/eng
3. Medical Device Innovation Consortium Real-World Clinical Evidence Generation Advancing Regulatory Science and Patient Access for In Vitro Diagnostics (IVDs) [cited 2024 Apr 30]. Available from: https://mdic.org/wp-content/uploads/2020/08/MDIC-2020-IVD-RWE-Framework_FINAL.pdf. Accessed April 29, 2024.
4. McDermott O, Kearney B. The value of using real-world evidence as a source of clinical evidence in the European medical device regulations: a mixed methods study. Expert Rev Med Devices. 2024;21(1-2):149-163. doi:10.1080/17434440.2023.2291454.
5. Wang C, Rosner GL, Bao T, et al. Leveraging real-world evidence for determining performance goals for medical device studies. Stat Med. 2021;40(29):6577-89. doi:10.1002/sim.9199.
6. Shi L, Xuan D, Jakovljevic M. A review on the evolving environment of medical device real-world evidence regulation on market access in the USA. Cost Eff Resour Alloc. 2024;22(1):75. doi:10.1186/s12962-024-00582-9.
7. Baumfeld Andre E, Gee M, Magnus C, et al. The Open Hand Initiative: Facilitating the use of real-world evidence in regulatory submissions through collaboration and transparency. Clin Pharmacol Ther. 2025;117(4):1072-77. doi:10.1002/cpt.3539.

Author information

Crispin Bennett is a Senior Medical Writer at Trilogy Writing & Consulting with extensive experience in the clinical development of both medical devices and pharmaceuticals. He has extensive experience in authoring regulatory documentation for combination products, bridging device and drug development. His expertise includes a varied set of documents across global regulatory environments.

Heidi Chapman is a Medical Writing Manager at Trilogy Writing & Consulting with broad experience in providing medical writing support to clinical research organisations, pharmaceutical companies, and medical device manufacturers. She has primarily written CEPs and CERs, developed templates, and written literature search protocols for drug-device combination products.

Laura C. Collada Ali is a Senior Medical Writing Manager at Trilogy Writing & Consulting. She has over 25 years of experience writing, editing, and managing medical and scientific documentation for diverse audiences. She is a member of EMWA's Executive Committee and holds multiple certificates in linguistics, pharmacology, medical devices and clinical research.

My First Medical Writing

SECTION EDITOR

Evguenia Alechine
ealechine@gmail.com

Editorial

Welcome to a new release of *My First Medical Writing*. In this edition, we bring you a contribution by Janaine Prata de Oliveira. As a former dentist with a PhD in Pharmacology, Janaine has worked throughout her scientific career on pain and inflammation research across Brazil, Belgium, and the USA. She is passionate about turning complex scientific and medical information into clear and relatable messages for a diverse audience, including children. Currently, she works as a post-doctoral researcher at St Louis University (USA) and as a freelance medical writer. You can find more information about her work at https://janawriteshealth.com. This article clearly shows her drive to communicate medicine and science-related topics to a broader audience. I hope you enjoy this read!

Evguenia

Gut feeling and chronic pain: Revisiting the microbiota connection

Janaine Prata de Oliveira
Freelance Medical Writer
St Louis, Missouri, USA
janaineprata@gmail.com

doi: 10.56012/bxad7017

Fibromyalgia, migraine, rheumatoid arthritis, irritable bowel disease (IBD), and neuropathic pain: Those conditions have one thing in common: chronic pain. Chronic pain is defined as pain that persists for three months or longer.[1] It affects nearly 20% of the global population and significantly impacts physical, emotional, and social well-being.[1,2]

Although chronic pain is a prevalent condition, treating it remains a challenge for patients and healthcare providers. Recent studies have shown a potential new player in this scenario: the gut microbiota.

The gut microbiota comprises thousands of microorganisms that perform essential functions beyond digestion. Emerging evidence highlights the role of the gut–brain axis in chronic pain, particularly through activation of immune cells and neuroinflammation. In this article, I explore how this interaction happens and how gut modulation might offer a potential strategy for chronic pain management. Let's dive into this microscopic world!

The gut microbiota: Beyond digestion

The gut microbiota is a microscopic world composed of over 1,000 microbial species and over 7,000 strains, primarily bacteria. In a balanced state (called homeostasis), the intestinal microbiota helps food digestion and immune regulation, as well as overall health.[3] On the other hand, when this balance is disrupted, the microbiome changes its composition, leading to dysbiosis. This condition can disturb the immune system and lead to chronic inflammation.[3-5]

In the past few years, the interaction between microbiota and the immune system has been investigated in many neurological and autoimmune conditions, such as Alzheimer's disease, multiple sclerosis, and chronic pain.[3] But how can small creatures make such an impact?

In the gut, microbial metabolism produces products such as lipopolysaccharides (LPS) and short-chain fatty acids (SCFAs). These microbial products can act as pro-inflammatory substances, activating immune cells that release cytokines and chemokines. Over time, this ongoing process exacerbates inflammation and can damage the intestinal barrier, a condition often referred to as "leaky gut". When the barrier is compromised, microbial products normally confined to the gut can cross into the bloodstream and reach distant organs, including the nervous system.[4-5]

Gut microbiota and chronic pain regulation

Gut microbes and their derivatives may contribute to peripheral sensitisation (pain) by directly activating specialised neurons in the peripheral nervous fibers (e.g., nerves in the skin, muscles, joints), called nociceptive neurons. Those gut microbes may also indirectly contribute to peripheral sensitisation by activating immune cells, and release of pro-inflammatory substances, such as IL-1β and TNF-β.[5,6]

> Importantly, recent clinical studies suggest that communication between the gut microbiota and microglia plays a key role in chronic pain development.

In the central mechanism, the gut microbiota interacts with the central nervous system, forming the gut-brain axis. Those microbial products (LPS and SCFAs) cross the blood-brain barrier that protects the brain from invaders. Then, they interact and activate the glial cells (microglia and astrocytes) and infiltrate immune cells. Those cells release additional inflammatory mediators and interact with neurons, causing neuroinflammation and exacerbating pain sensitisation.[5,7,8]

Importantly, recent clinical studies suggest that communication between the gut microbiota and microglia plays a key role in chronic pain development.[4,5,9-11] Therefore, restoration of gut microbiota homeostasis can reduce microglia activation, hence neuroinflammation and chronic pain. [4,5]

Microbiota-based therapies

Many chronic pain conditions present an imbalance in the gut microbiota, often marked by an increase of obligately or facultatively anaerobic and aerobic bacteria, such as *Clostridiaceae*, *Enterobacteriaceae*, and *Proteobacteria*.[4,5] At the same time, there is a reduction in the bacteria population that live in harmony with the body (also called commensal bacteria), disrupting the microbial ecosystem. In addition, pain may affect the microbiota profile across different types of painful conditions.[5,6] Thereby, some alternatives have been proposed to reestablish gut microbiota equilibrium.

Probiotics

Probiotics are active microorganisms that promote health benefits to the host, including immune modulation.[12] For instance, a study showed that probiotic supplementation of Lactobacilli, *Bifidobacterium,* and *Saccharomyces* for eight weeks reduced pain and improved sleep quality (i.e., getting uninterrupted and refreshing sleep), depression, and anxiety in patients with fibromyalgia.[13] Another study showed that supplementation of *Lactobacillus* for four weeks reduced pain in children with inflammatory bowel syndrome.[14] The positive effect of probiotic supplementation has also been confirmed in migraine after use of *Lactobacilli*, *Bifidobacteria,* and *Streptococcus* for ten weeks.[15]

Fecal microbiota transplantation

Fecal microbiota transplantation (FMT) transfers a small stool from a healthy donor into the gastrointestinal tract of a patient to modify intestinal flora and restore gut homeostasis.[16] In patients suffering from fibromyalgia, FMT reduced widespread pain, anxiety, and depression, and improved sleep quality after 3, 6, and 12 months after intervention.[17] FMT transplantation also reduced abdominal pain, discomfort, and severity in patients with inflammatory bowel syndrome after 12 weeks of the procedure; however, these effects were reduced over 1 year.[18]

Perspectives

The egg or the chicken? Which comes first, the neuroinflammation and chronic pain, or the dysbiosis? This is one of the questions that scientists are still uncovering. The relationship between gut microbiome and chronic pain is undeniable; however, we need to understand more about multiple microbiota profiles across different types of chronic pain and their clinical implications.

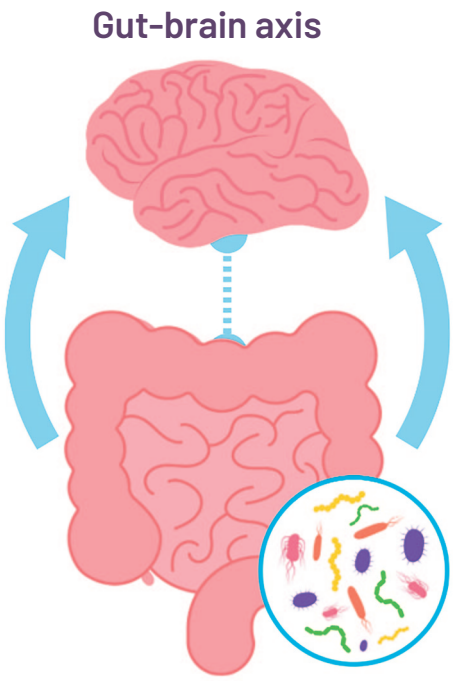
For now, a multidisciplinary approach remains essential. Additionally, patients should actively engage in their treatment plans to effectively manage chronic pain: following a balanced diet, practicing regular physical activity, working on mental health, and adhering to healthcare guidance. There is no simple solution. But the gut microbiota may offer new hope for chronic pain treatment.

Disclaimers

The opinions expressed in this article are the author's own and are not necessarily shared by their employers or EMWA.

Disclosures and conflicts of interest

The author declares no conflicts of interest.



References

1. Treede RD, Rief W, Barke A, et al. Chronic pain as a symptom or a disease: the IASP Classification of Chronic Pain for the International Classification of Diseases (ICD-11). Pain. 2019;160(1):19-27. doi:10.1097/j.pain.0000000000001384
2. Zimmer Z, Fraser K, Grol-Prokopczyk H, et al. A global study of pain prevalence across 52 countries: examining the role of country-level contextual factors. Pain. 2022;163(9):1740-50. doi:10.1097/j.pain.0000000000002557
3. Lloyd-Price J, Abu-Ali G, Huttenhower C. The healthy human microbiome. Genome Med. 2016;8(1). doi:10.1186/s13073-016-0307-y
4. Ustianowska K, Ustianowski Ł, Machaj F, et al. The role of the human microbiome in the pathogenesis of pain. Int J Mol Sci. 2022;23(21):13267. doi:10.1007/s12035-024-04663-x
5. Lou L, Zhou L, Wang Y. Gut microbiota: a modulator and therapeutic target for chronic pain. Mol Neurobiol. 2025 May;62(5):5875-90. doi:10.1007/s12035-024-04663-x.
6. Guo R, Chen LH, Xing C, et al. Pain regulation by gut microbiota: molecular mechanisms and therapeutic potential. Br J Anaesth. 2019;123(5):637-654. doi:10.1016/j.bja.2019.07.026
7. Ho T, Elma Ö, Kocanda L, et al. The brain-gut axis and chronic pain: mechanisms and therapeutic opportunities. Front Neurosci. 2025;19. doi:10.3389/fnins.2025.154599
8. Morreale C, Bresesti I, Bosi A, et al. Microbiota and pain: save your gut feeling. Cells. 2022 Mar;11(6):971. doi:10.3390/cells11060971
9. Boer CG, Radjabzadeh D, Medina-Gomez C, et al. Intestinal microbiome composition and its relation to joint pain and inflammation. Nat Commun. 2019;10(1):4881. doi:10.1038/s41467-019-12873-4
10. Kopchak OO, Hrytsenko OY, Pulyk OR. Peculiarities of the gut microbiota in patients with migraine comparing to healthy individuals. Wiad Lek. 2022;75(9 pt 2):2218-2221. doi:10.36740/WLek202209207
11. Chen J, Wang Q, Wang A, et al. Structural

Illustration: Janaine Prata de Oliveira

and functional characterization of the gut microbiota in elderly women with migraine. Front Cell Infect Microbiol. 2020;9:470. doi:10.3389/fcimb.2019.00470

12. Butel MJ. Probiotics, gut microbiota and health. Med Mal Infect. 2014;44(1):1-8. doi:10.1016/j.medmal.2013.10.002
13. Aslan ÇİN NN, Açik M, TertemİZ OF, et al. Effect of prebiotic and probiotic supplementation on reduced pain in patients with fibromyalgia syndrome: a double-blind, placebo-controlled randomized clinical trial. Psychol Health Med. 2024;29(3):528-541. doi:10.1080/13548506.2023.2216464
14. Weizman Z, Abu-Abed J, Binsztok M. Lactobacillus reuteri DSM 17938 for the management of functional abdominal pain in childhood: a randomized, double-blind, placebo-controlled trial. J Pediatr. 2016;174:160-64.e1. doi:10.1016/j.jpeds.2016.04.003
15. Martami F, Togha M, Seifishahpar M, et al. The effects of a multispecies probiotic supplement on inflammatory markers and episodic and chronic migraine characteristics: A randomized double-blind controlled trial. Cephalalgia. 2019;39(7):841-853. doi:10.1177/0333102418820102
16. Bowman KA, Broussard EK, Surawicz CM. Fecal microbiota transplantation: current clinical efficacy and future prospects. Clin Exp Gastroenterol. 2015;8:285-91. doi:10.2147/CEG.S61305
17. Fang H, Hou Q, Zhang W, et al. Fecal microbiota transplantation improves clinical symptoms of fibromyalgia: an open-label, randomized, nonplacebo-controlled study. J Pain. 2024;25(9):104535. doi: 10.1016/j.jpain.2024.104535
18. Holvoet T, Joossens M, Vázquez-Castellanos JF, et al. Fecal microbiota transplantation reduces symptoms in some patients with irritable bowel syndrome with predominant abdominal bloating: short- and long-term results from a placebo-controlled randomized trial. Gastroenterology. 2021;160(1):145-57.e8. doi:10.1053/j.gastro.2020.07.013.

Janaine Prata de Oliveira

CONTACT US

If you have ideas for themes or would like to discuss any other issues, please write to mew@emwa.org.

Upcoming issues of **Medical Writing**

December 2025

Safety Writing

As the regulatory landscape continues to evolve, the importance of precise and thorough safety reporting has never been more critical. This issue will provide insights into the latest methodologies, best practices, and innovative approaches that are shaping the future of safety writing. The issue will feature articles on the development and submission of safety data, offering expert guidance on handling complex safety data.

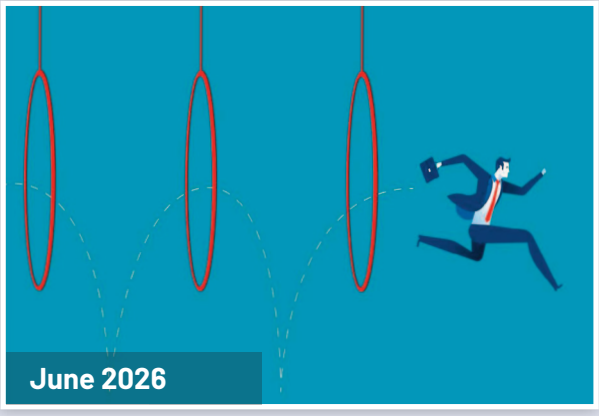
Guest Editors: Iva Cvetkovic and Pavle Simeunovic
The deadline for feature articles has now passed.

March 2026

Medical Writing Careers

In this issue, we are exploring careers within medical writing, including getting that first role, internship programmes, different types of medical writing, and how a medical writing career may not go in the direction you expected. We will look at how to become more specialised, how to move into management, and potential opportunities after you stop writing.

Guest Editors: Andrew Balkin and Jules Kovacevic
The deadline for feature articles is December 1, 2025.



June 2026

Regulatory Submissions

This issue explores the modern regulatory submission landscape, reflecting the innovative strategies being employed to navigate it. The issue will underscore a pivotal shift in the industry: from viewing submissions as a final administrative step to embracing them as a dynamic, strategic process integral to product development and lifecycle management. This issue offers insights for medical writers working in pharmaceuticals/biotech, medical devices, and veterinary medicine industries.

Guest Editors: Clare Chang and Annie Josie
The deadline for feature articles is March 1, 2026.

September 2026

Publications: Integrity, Ethics, and Peer Review

This issue will cover the latest approaches to publishing ethics implementation, technology-enabled fraud detection, and institutional integrity transformation. It will also feature articles from thought leaders at major publishers, pharmaceutical companies, and technology firms on preventing integrity violations through proactive culture change rather than reactive compliance.

Guest Editors: Martin Delahunty and Slávka Baronikova
The deadline for feature articles is June 1, 2026.

http://journal.emwa.org/