Volume 29 Number 2
June 2020

Medical Writing

The data economy

Also in this issue...

- Digital identifiers in scientific publishing and e-health
- Master trial documents for increased efficiency and scientific integrity

EMWA EUROPEAN MEDICAL WRITERS ASSOCIATION

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.

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

Editorial: The data economy 1
Raquel Billiones, Sam Hamilton

The data economy: A glossary 3
Raquel Billiones

Presidents' Messages 8
Beatrix Doerr, Barbara Grossman

EMWA News 10

Viewpoint 14
Rush to publication – What do we have to lose?
Art Gertel

Features

The value of registry data in the clinical evaluation of medical devices 16
Kelly Goodwin Burri, Adrian Spoerri

Big data in clinical research: Present and future 22
Hyunjoo Kim, Zuo Yen Lee, Chow Lih Yew

Leveraging standardised data in response to the novel coronavirus outbreak 32
Derk Arts, Lauren McGill

Biostatistics, data management, and medical writing: A multidisciplinary approach to the development of the CTD integrated summaries 36
Shiri Diskin, Tanya du Plessis, Ofra Barnett-Griness

The new value of clinical data in Europe 42
Jasminka Roth

Unlocking the potential of patient data through responsible sharing – has anyone seen my keys? 46
Patrick Cullinan, Liz Roberts

COVID-19 contagion, information, and misinformation 52
Melvin Sanicas

Processing health-related data for scientific research: Is consent an appropriate legitimate ground? 56
An Vijverman

An interview with Cathal Gallagher of EMA's Technical Anonymisation Group 58
Sam Hamilton, Cathal Gallagher

A role for medical writers in overcoming commonly held misconceptions around FAIR data 60
Erik Schultes

MyData: Applying human-centric principles to health data 64
MyData Global Author Group

Thriving in the data economy as data-fluent PhD graduates 70
Menorca Chaturvedi

Digital identifiers in scientific publishing and e-health 74
Hamid Reza Khedmatgozar

Special Contributions

Winners of the Geoff Hall Scholarship Essay Competition 79

A poem for my client 103

Regular Features

News from the EMA 84

Regulatory Matters 88
- Master trial documents for increased efficiency and scientific integrity

Regulatory Public Disclosure 90

Medical Devices 92
- Medical writers under lockdown

Getting Your Foot in the Door 94
- Big feet make for a great door stopper
- My journey into the industry

Journal Watch 97

Veterinary Medical Writing 100
- Research in veterinary medicine takes off!
- Regulatory in veterinary medicine takes off!
- A tale of "One Health" from Japan

Good Writing Practice 104

Out on Our Own 107
- Switching careers: My path to medical writing and freelancing

Volume 29 Number 2 | June 2020

Medical Writing

The data economy

Data are economic assets that power the so-called fourth industrial revolution. The healthcare industry is at the forefront of this "data economy". Medical writers should understand how to use these data appropriately and responsibly. This issue of *Medical Writing* is dedicated to our vital place in the data economy. A glossary of relevant data-related terms is provided on p. 3.

Medical writers and communicators support data generation by writing and reviewing documents that report clinical trial data and the methods used to collect them. Real-world data (RWD) are also increasingly collected outside the controlled environment of clinical trials through mobile devices and patient registries, and the US FDA and EMA encourage their use as evidence to support efficacy and safety of health products. Leveraging big RWD to develop therapeutics is challenging for the medical writing community, which is more accustomed to smaller clinical trials databases; we need to learn about the reliability of RWD and how they can be used. In their article on p.16, **Kelly Goodwin Burri** and **Adrian Spoerri** describe how health registries are used to collect RWD for the clinical evaluation of medical devices.

The healthcare industry is at the forefront of the data economy ... As medical writers, we need to collect, use, and share data wisely and responsibly.

Regulatory authorities are aware of the problem of lack of generalisability of clinical trials and are implementing strategies to support the clear message of ICH E8(R1) General Considerations for Clinical Trials, 08 May 2019 that clinical trial protocols need to explore non-conventional data sources including "big data". The US FDA recognises that to make clinical trials more generalisable, they must suggest trial designs that better reflect the populations that medicines serve. The June 2019 Draft FDA Guidance for Industry "Enhancing the Diversity of Clinical Trial Populations – Eligibility Criteria, Enrollment Practices, and Trial Designs" addresses the enrichment of clinical trials. Among the many suggested strategies is early engagement with patient advocacy groups to elicit their suggestions for clinical trial design. In June 2019, the EMA announced a collaboration with European primary care doctors to gather RWD on medicines typically used in the primary care setting to strengthen the pharmacovigilance research base. Further, in January 2020, the Heads of Medicines Agency-EMA Big Data Task Force announced ambitious plans to unlock big data for public health benefit. Their 10 recommendations are topped by a plan to "…deliver a sustainable platform to access and analyse healthcare data from across the EU (Data Analysis and Real World Interrogation Network-DARWIN)". Further, the EMA's Information Management Strategy 2020 to 2022 prioritises "…dialogue with stakeholders on big data, real world data and artificial intelligence to ensure EMA is informed on opportunities for collaboration and able to facilitate data access and analytics".

So how might all this affect the daily work of medical writers? At the very least, we need to rethink conventional clinical trial design. In their article on p. 22, **Hyunjoo Kim and colleagues** explore this idea as well as the impact on protocol and clinical study report authoring, and they provide general insights on how big data might change the clinical-regulatory medical writing landscape.

In the midst of the COVID-19 global outbreak, effective real-time data collection is crucial for the healthcare system to prepare and respond to unfolding events. **Derk Arts and colleagues** discuss this problem and present solutions on p. 32.

The pandemic has created significant disruptions in protocol-specified procedures, data collection, analysis, and reporting for ongoing clinical trials. Regulatory bodies were quick to react and released emergency guidance documents in March 2020, such as the FDA Guidance on Conduct of Clinical Trials of Medical Products during COVID-19 Pandemic and the EMA Guidance on the Management of Clinical Trials during the COVID-19 (Coronavirus) Pandemic, which were subsequently updated as events

GUEST EDITORS

Raquel Billiones
medical.writing@billiones.biz

Sam Hamilton
sam@samhamiltonmwservices.co.uk

About the Guest Editors

Dr Sam Hamilton is a post-doctoral virologist, currently Global Head of Medical and Regulatory Writing and Public Disclosure for the CRO, Clinipace. With 26 years in clinical and regulatory medical writing roles in the pharmaceutical industry, Sam is independently responsible for her wider professional interests. Sam's interest in public disclosure of clinical-regulatory documents has grown since chairing the EMWA-AMWA group who delivered open-access www.core-reference.org in May 2016. A long-time supporter of EMWA, Sam has served in various roles, notably as Freelance Advocate; Editorial Board member for *Medical Writing*; Workshop Leader; Expert Seminar Series Chair; and Vice President and President of the Executive Committee. Sam was elected an EMWA Lifetime Fellow in 2018 for her services to the association, and is currently *Medical Writing* Section Editor for the "Regulatory Public Disclosure" Section and on the Advisory Panel of the Regulatory Public Disclosure Special Interest Group.

Raquel Billiones, PhD Biology, has been a regulatory writer for more than 14 years, covering both pharmaceuticals and medical devices. Her core competencies include clinical trials and submissions documents, data disclosure and protection, and project and people management. Over the years, she took on a wide range of industry positions, as freelancer, employed regulatory writer, and as head of medical writing departments in the CRO and big pharma settings. Raquel is an active EMWA member, serving in various roles, including as Executive Committee member (2015–2017), journal Associate Editor, workshop leader, EPDC member, Medical Device SIG chair, and Sustainability SIG co-founder.

unfolded. We can be sure that the data generated by the COVID-19 pandemic will provide a rich and extensive big data archive.

As medical writers, we consume health data: we use data to communicate study results to the authorities, the scientific community, and the public. And data begets data as we process and analyse data collected and use the results of those analyses to generate more data and move medical and scientific research forward. **Shiri Diskin and colleagues** describe the multidisciplinary approach of integrating different datasets into cohesive summaries that support market authorisation of medicinal products (p. 36), and on p. 42, **Jasminka Roth** expounds the merits of meta-analysis of multiple clinical studies to support a medical device's certification for market access.

In addition to consuming health data, in the era of data transparency and disclosure, we help disseminate data. But data sharing comes with the responsibility to protect the privacy of the individual data subject. Previously a reluctant player, the pharmaceutical industry is now taking a lead role in sharing data proactively and responsibly. There are myriad ways and platforms to share clinical data, as disclosure experts **Patrick Cullinan and Liz Roberts** present on p. 46.

On the social media front, misinformation and disinformation are rampant, highlighting the importance of reliable data sources. In the context of the COVID-19 outbreak, tropical disease expert **Melvin Sanicas** shares his views on responsible social media sharing of health information and the data sources that he uses on p. 52.

The GDPR aspect of big data sharing is explored in two regulatory-related articles that begin on p. 56. **Sam Hamilton** engages two experts: **An Vijverman,** a Brussels-based lawyer and expert in the legalities of health data processing, and **Cathal Gallagher** of EMA's Technical Anonymisation Group.

We need to collect, use, and share data wisely. The data economy is dominated by big data, characterised by high **v**olume, high **v**elocity, and wide **v**ariety. However, big health data exist in different structures and are stored in different repositories. Despite the common use of computers and artificial intelligence in healthcare, the three "Vs" of big data still present a major challenge. According to data scientists, for data to be used effectively, it must be Findable, Accessible, Interoperable, and Reusable (FAIR). **Erik Schultes** explains the FAIR data guiding principles in health and medicine on p. 60, using the COVID-19 outbreak as an example.

So who owns all the health data collected? Are these data given freely by individual subjects? The **MyData group** presents a plea for human centric control over health data on p. 64.

Finally, how does the future of medical writing and communications look in the data economy? **Menorca Chaturvedi**, recipient of the EMWA Geoff Hall Scholarship explains on p. 70 how the workplace is dominated by "hybrid jobs" that require data literacy and communication skills.

We want to end our introduction to this complex topic with a big Thank You to all our contributors. To our readership, we hope that you learn as much as us in reading about big data and the data economy as we learned in putting this issue together. Enjoy!

The data economy
A glossary

The data economy comes with its own terminologies and buzzwords (Table 1), stakeholders (Table 2), and activities (Table 3). This glossary aims to help readers navigate this data-driven environment.

Table 1. Key terms

Term	Definition	Source
Personal data	• Any information relating to an identified or identifiable living individual ("natural person"; "data subject"); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data. • Different pieces of information, which collected together can lead to the identification of a particular person, also constitute personal data.	GDPR Article 4 (1) https://ec.europa.eu/info/law/law-topic/data-protection/reform/what-personal-data_en
Genetic data	• Personal data relating to the inherited or acquired genetic characteristics of a natural person which give unique information about the physiology or the health of that natural person and which result, in particular, from an analysis of a biological sample from the natural person in question.	GDPR Article 4 (13)
Biometric data	• Personal data resulting from specific technical processing relating to the physical, physiological or behavioural characteristics of a natural person, which allow or confirm the unique identification of that natural person, such as facial images or dactyloscopic (e.g., fingerprints) data.	GDPR Article 4 (14)
Sensitive personal data	• Personal data which are, by their nature, particularly sensitive in relation to fundamental rights and freedoms and the context of their processing could create significant risks to those fundamental rights and freedoms. • Example of sensitive data are race or ethnic origin, political opinions, religion or beliefs, trade union membership, genetic data, data on health status or sexual orientation.	GDPR Preamble 10, 51, 71
Health data or data concerning health	• Personal data related to the physical or mental health of a natural person, including the provision of health care services, which reveal information about his or her health status.	GDPR Article 4 (15)
Protected personal data (PPD)	• Any information relating to an identified or identifiable natural person ("data subject"); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity.	EMA Policy 0070
Individual patient data (IPD)	• Individual data separately recorded for each participant in a clinical study.	EMA Policy 0070
Aggregated data	• In the context of clinical studies, represent statistical data about several individuals that has been combined to show general trends or values without identifying individuals within the data.	EMA Policy 0070

Term	Definition	Source
Anonymous data	• Information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable, also called "de-identified data". • Anonymised data is not considered personal data and is out of scope of GDPR. For data to be truly anonymised, the anonymisation must be irreversible. • Clinical trial data are not fully anonymised.	GDPR Preamble 26 EMA Policy 0070
Pseudonymous data	• Processing of personal data in such a manner that the data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person. Personal data that has been de-identified, encrypted or pseudonymised but can be used to re-identify a person remains personal data and falls within the scope of the GDPR. • Clinical trial data are pseudonymised data.	EMA Policy 0070
Identifier	• Information that can directly or indirectly identify a data subject.	EMA Policy 0070
Direct identifiers	• Data elements that permit direct recognition or communication with the corresponding individuals. • Examples: name, email, phone number, address, patient identification number.	EMA Policy 0070
Indirect ("quasi") identifiers	• Data elements representing an individual's background information that can indirectly identify data subjects. • Examples: demographics, characteristics, attributes, socio-economic information.	EMA Policy 0070
Big data	• Term applied to data sets whose size or type is beyond the ability of traditional relational databases to capture, manage and process the data with low latency. Big data has one or more of the following characteristics "the three Vs"): high Volume, high Velocity or wide Variety. Artificial intelligence (AI), mobile, social and the Internet of Things (IoT) are driving data complexity through new forms and sources of data. For example, big data comes from sensors, devices, video/audio, networks, log files, transactional applications, web, and social media – much of it generated in real time and at a very large scale.	https://www.ibm.com/analytics/hadoop/big-data-analytics
Real world data (RWD)	• Data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources.	https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence
Real world evidence (RWE)	• Clinical evidence regarding the usage and potential benefits or risks of a medical product derived from analysis of RWD. RWE can be generated by different study designs or analyses, including but not limited to, randomised trials, including large simple trials, pragmatic trials, and observational studies (prospective and/or retrospective).	https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence

Table 2. Key stakeholders

Term	Definition	Source
Data subject	• Any natural ("living") person whose personal data is being processed; see definition of data processing.	GDPR
Identifiable natural person	• One who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.	GDPR Article 4 (1)
Data controller	• Natural or legal person, public authority, agency or any other body which alone or jointly with others determines the purposes, conditions and means of the processing of personal data.	GDPR Article 4 (7)
Data processor	• A natural or legal person, public authority, agency or any other body which processes personal data on behalf of the controller.	GDPR Article 4 (8)
Data recipient	• A natural or legal person, public authority, agency or another body, to which the personal data are disclosed, whether a third party or not. However, public authorities which may receive personal data in the framework of a particular inquiry in accordance with Union or Member State law shall not be regarded as recipients; the processing of those data by those public authorities shall be in compliance with the applicable data protection rules according to the purposes of the processing.	GDPR Article 4 (9)
Data exporter	• A natural or legal person, public authority, agency or another body who transfers personal data from the EU/EEA to a non-EEA country or international organisation. • In the context of international cooperation due to COVID-19, international transfers of health data for the purpose of scientific research outside of the EEA are allowed under certain conditions. The exporter must meet GDPR requirements for data transfers.	European Data Protection Board Guidelines 03/2020 on the processing of data concerning health for the purpose of scientific research in the context of the COVID-19 outbreak
Data scientist	• A professional engaged in data science. A data scientist requires an integrated skillset spanning mathematics, machine learning, AI, statistics, databases, and optimisation.	Dhar V. Data Science and Prediction Communications of the ACM, December 2013, Vol. 56 No. 12, pp. 64–73 10.1145/2500499
Data protection authorities (DPA)	• Independent public authorities in each EU member state that supervise, through investigative and corrective powers, the application of the data protection law.	GDPR

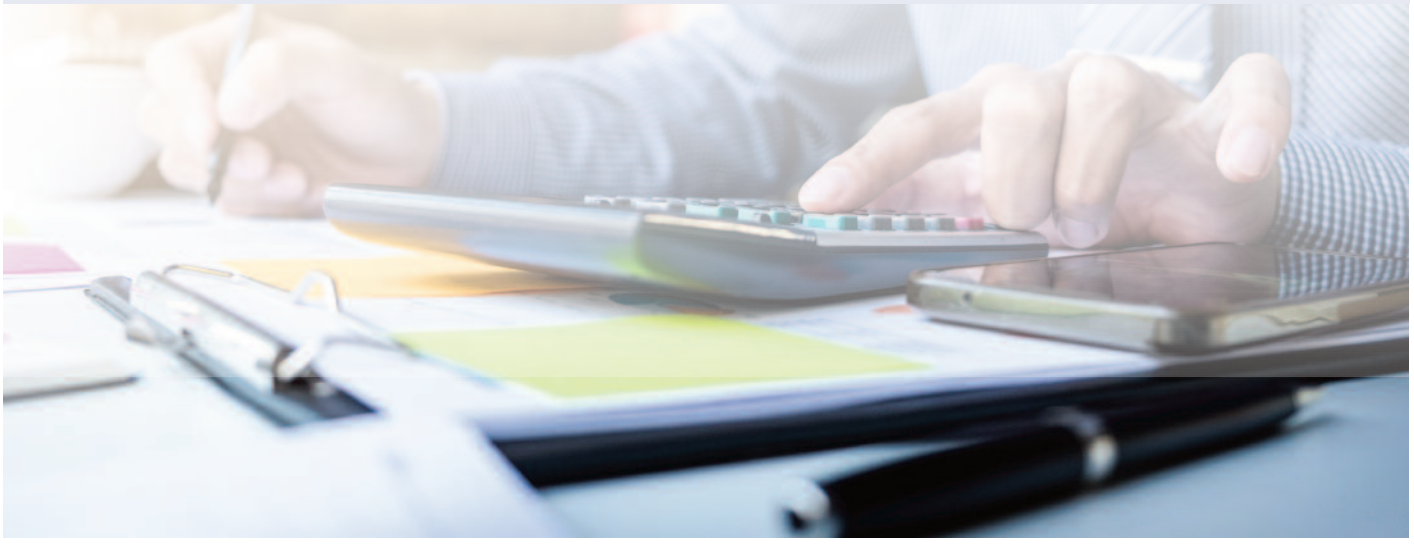


Table 3. Key activities

Term	Definition	Source
Data science	• The study of the generalisable extraction of knowledge from data.	Dhar V. Data Science and Prediction Communications of the ACM, December 2013, Vol. 56 No. 12, pp. 64–73 10.1145/2500499
Data processing	• Processing covers a wide range of operations performed on personal data, including by manual or automated means. It includes the collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction of personal data. • Data collection, analysis, reporting, and sharing during a clinical trial constitute data processing.	https://ec.europa.eu/info/law/ law-topic/data-protection/ reform/what-constitutes-data-processing_en GDPR Article 4 (2)
Data sharing	• Making data available to others.	
Data transparency	• In the context of GDPR, personal data must be processed "lawfully, fairly and in **a transparent manner** in relation to the data subject." This covers the data subject's right to information (at the minimum) about data use, storage, access, and dissemination. • In the context of clinical research, data transparency is sharing of clinical research data to meet regulatory requirements and advance the generation of critical scientific knowledge.	GDPR Article 5 (1a) https://www.phusewiki.org/docs/ WorkingGroups/New%20 Template% 20Deliverables/Data%20Transparency/ Clinical%20Trial%20Transparency% 20and%20Disclosure-%20A%20 Global% 20View.pdf
Personal data protection	• The act of protecting the rights of data subjects.	GDPR
Anonymisation	• The process of rendering data into a form which does not identify individuals and where identification is not likely to take place. Anonymisation is irreversible.	EMA Policy 0070

Table 3 continued opposite

Term	Definition	Source
Pseudonymisation	• The process of replacing one attribute (typically a unique attribute) in a record by another. The natural person is may still be identified indirectly but pseudonymisation reduces the linkability of a dataset with the original identity of a data subject.	EMA Policy 0070
Re-identification or de-anonymisation	• The process of analysing data or combining it with other data with the result that individuals become identifiable.	EMA Policy 0070
Anonymisation techniques	• Techniques to mitigate risks of re-identification of the individual data subjects; Effective ness of these techniques are based on three criteria: singling out; linkability, and inference.	EMA Policy 0070
Proactive anonymisation (in the context of the CSR)	• Use of anonymisation techniques for generating anonymised datasets and generating another copy of the CSR using anonymised datasets. • If it is necessary to discuss any individual subject level information in text, consider using **proactively anonymised clinical document text and data presentations** that maintain data meaning, remain in context AND conform to **current** minimum standards for de-identifying data.	https://www.phusewiki.org/docs/Deliverables/Narratives%20Phuse%20Subgroup%20Writeup_Final_11.21%20(1).pdf https://www.core-reference. org/core-reference/
Retrospective anonymisation	• Generally, the use of redaction or masking to anonymise text and data; also known as reactive data anonymisation.	EMA Policy 0070 https://www.phusewiki.org/docs/Deliverables/Narratives%20Phuse%20Subgroup%20Writeup_Final_11.21%20(1).pdf
Big data analytics	• Use of advanced analytic techniques against very large, diverse data sets that include structured, semi-structured and unstructured data, from different sources, and in different sizes from terabytes to zettabytes.	https://www.ibm.com/analytics/hadoop/big-data-analytics

Raquel Billiones
medical.writing@billiones.biz

The data economy

Presidents' Messages

From EMWA President 2020–2021

We now look ahead and focus our efforts related to remote learning options.

When I had some spare time in January, I worked on this message as I expected April to be quite packed. Being written before the COVID-19 crisis, the message was positive in every aspect – it was about the progress of EMWA within the past year and what exciting times and new projects are to come. Well, this was in vain, as the circumstances have changed dramatically and nobody knows what is going to happen. Hopefully, things will have improved by the time this issue is published in June.

In response to the COVID-19 pandemic, we had to cancel our Spring Conference. Meanwhile, we also cancelled the face-to-face Autumn conference. This was a decision the Executive Committee did not take lightly and was made because our members' health and safety are the top priority. By making the decision early, we will be able to focus time and resources on planning for an exciting virtual conference in November.

Needless to say, the cancellation of the conferences is a major financial loss for EMWA. Luckily, we have been cautious in spending money during the past years and now have a robust financial reserve to be able to withstand it.

Being occupied most of the past weeks with disaster recovery and planning, we now look ahead and focus our efforts related to remote learning options:

- The first virtual Expert Seminar Series (June)
- Virtual round table discussions (starting by September)
- Virtual fall conference (November)

Let me take this opportunity to express my deepest gratitude to the webinar team, led by EMWA Professional Development Committee members Laura Collada Ali and Carolina Rojido, for all they have done during the past weeks. They navigated us through different platforms and formats, helped us prepare for the first virtual annual meeting and the first virtual Expert Seminar Series, and organised additional webinars. They took on these giant tasks despite the challenges of their daily lives and jobs.

Aside of these activities, we have released our EMWA member logo, which you can now download from the members-only area of the EMWA website (https://tinyurl.com/ydfowwtf). You can use it on your website and in your email signature, which can increase your market value by showing your membership in a professional organisation.

Let me end in thanking our former president Barbara Grossman, for all her help and work during the past years and for leading the Executive Committee and EMWA through the challenges of the early months of this year. Looking into the future, I am happy that Carola Krause will be supporting us as Vice President. Her technical knowledge and expertise as a digital native is particularly relevant in these turbulent times.

I deeply miss seeing you all in person. As someone who works from home, the conferences have always been a highlight of the year. I will truly miss meeting colleagues, friends, and new people from all over the world and visiting interesting locations. But let's stay "together apart" and make the best out of the situation.

My thoughts are with you and your families to stay safe and healthy during these challenging times.

Beatrix Doerr

From EMWA President 2019–2020

Beatrix Doerr, our new EMWA President, has neatly summarised the impact on "normal EMWA life" of the COVID-19 pandemic and the steps that the Executive Committee (EC) and the Education Committee are taking to bring EMWA to YOU – through different remote learning options. For example, as I write this, the first virtual Expert Seminar Series has just been announced.

What a contrast to 2019 with the highs of two well-attended EMWA conferences in Vienna and Malmö and the many EMWA achievements between times. In this message, I'd like to take the opportunity to thank a few people, for example: the wonderfully supportive EC, the Head Office team (led by Lynne Fletcher, Claire Whittingham and Lisa Wilson), and YOU – EMWA's many volunteers. Thank you. I thoroughly enjoyed working with all of you, especially as you made even the most challenging tasks doable.

In particular, I'd like to thank:

- Carolina Rojido and Laura Collada Ali with their Webinar Team, who organise a great variety of presentations – with thanks also to some excellent presenters! If you didn't manage to catch past webinars on the day, you can view them through the website: go to https://www.emwa.org/training/emwa-webinars-programme-2020/
- The Education Committee led by Marian Hodges, who worked with EMWA Head Office to plan and implement a diverse offering of workshops at both the May and November 2019 conferences – a mammoth logistical challenge.
- Those talented EMWA members who are also linguists; because of their efforts, EMWA has provided translations of the following Joint Position Statements:
  - The Role of Professional Medical Writers
  - Predatory Publishing

  The translations help to spread awareness of the responsibilities of medical communicators among non-English speakers.
- Our team of Ambassadors, led by Abe Shevack, who represent EMWA at a variety of events and last year participated in a video for medics considering a move to medical writing; see https://www.emwa.org/about-us/ambassadors-programme/
- The Social Media team led by Maria João Almeida, who effectively communicate to the "world outside EMWA" as well as to EMWA members.
- Phil Leventhal (Editor-in-Chief), his Editorial Team, and Vicki White (Managing Editor), who together with several outstanding guest editors and contributors, ensure that the EMWA journal, *Medical Writing*, is a "good read".
- Diarmuid De Faoite and the Website Team, who together with IT support at Head Office ensure that the EMWA website is up-to-date and readily searchable.

Although we won't meet this year, I hope to see many of you at future conferences. In the meantime, STAY HEALTHY, and I wish all of you well – you, your families, and your friends. Please join me in wishing Beatrix and the new EC every success.

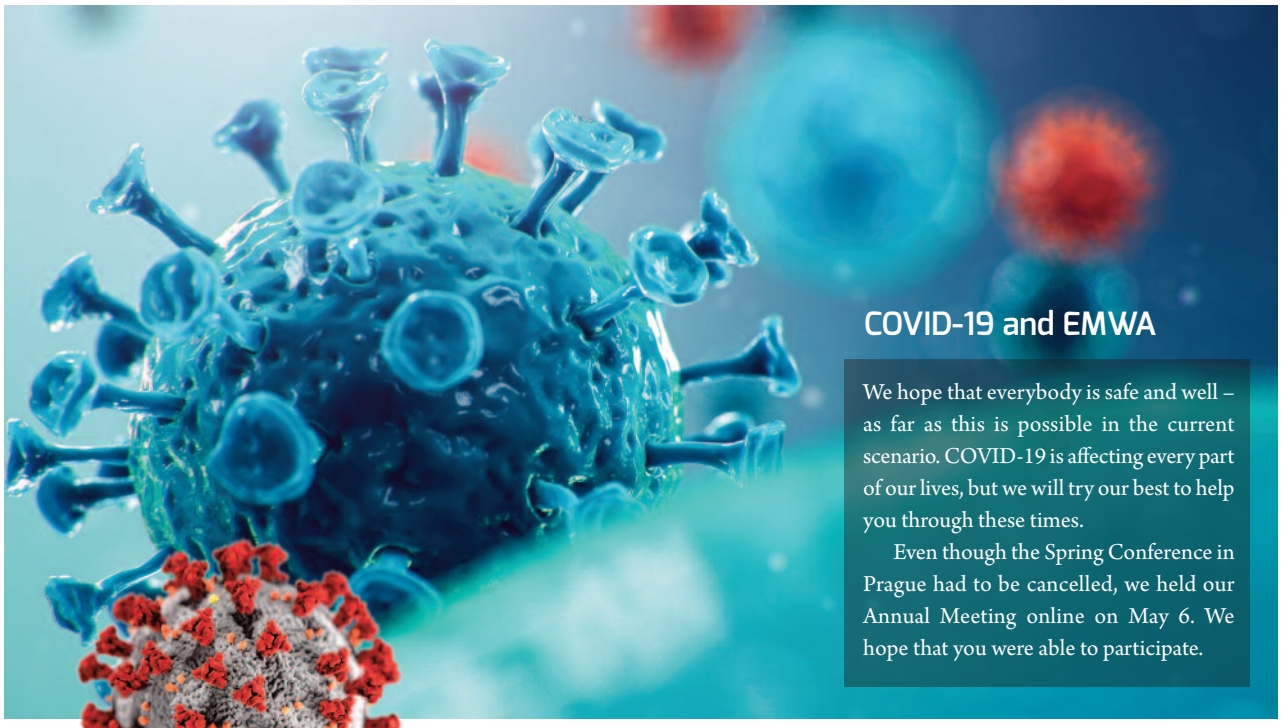
Barbara Grossman

Thank you. I thoroughly enjoyed working with all of you, especially as you made even the most challenging tasks doable.

EMWA News

SECTION EDITOR

Somsuvro Basu
basu.somsuvro@gmail.com



COVID-19 and EMWA

We hope that everybody is safe and well – as far as this is possible in the current scenario. COVID-19 is affecting every part of our lives, but we will try our best to help you through these times.

Even though the Spring Conference in Prague had to be cancelled, we held our Annual Meeting online on May 6. We hope that you were able to participate.

EMWA webinars

In these unprecedented times, we suggest that you use the wealth of our training material in the Webinars Programme Archive, which we plan to make more user-friendly by preparing search categories.

Given the exceptional circumstances due to COVID-19, we presented an additional webinar: "Lifestyle choices for medical writers" by Carolina Rojido, on April 23**.** You can listen to the recording in the Webinars Archive: https://members.emwa.org/EMWA/ Member_Area/zEMWA_Webinars_Programme_-_Archive.aspx

We encourage you to explore the varieties of webinars stored in our archive, spanning from career development, various tools of medical writing, to invaluable advice from seasoned medical writers.

Upcoming EMWA webinars:

- September 2020 *(exact date to be confirmed)*
  First experiences in the writing of Summaries of Safety and Clinical Performance (SSCPs) for medical devices
  Laura C. Collada Ali, Medical Writing Consultant & Helen Frampton, Oxford Medical Writing
- October 2020 *(exact date to be confirmed)*
  Transitioning from Medical Translation to Medical Writing
  Laura C. Collada Ali, Medical Writing Consultant & Paz Gómez-Polledo, freelance Medical Writing & Translation Consultant

Translation of the Joint Position Statement on Predatory Publishing into Polish and Swedish

The American Medical Writers Association (AMWA), the European Medical Writers Association (EMWA) and the International Society for Medical Publication Professionals (ISMPP) recognise the challenges to scientific publishing being posed by predatory journals and their publishers, which employ practices undermining the quality, integrity, and reliability of published scientific research. The joint position statement (JPS) complements several other sets of guidelines that have helped define the characteristics of a predatory journal.

By joining with AMWA and ISMPP in both developing and publicising the Joint Position Statement on Predatory Publishing, EMWA is providing a valuable service to publication professionals around the world by enabling them to more easily read, understand, and apply the principles of this JPS.

To raise awareness among non-English speakers about the responsibilities of medical writers and publication professionals concerning this significant issue, EMWA has initiated the translation of this statement into European languages.

We are proud to announce the posting of the first JPS translation into Polish by Maria Kołtowska-Häggström, Dorota Szymańska, Jacek Bil, and Olga Mozenska. The translation is available on the EMWA website: https://www.emwa.org/about-us/position-statements/joint-position-statement-on-predatory-publishing/polish/.

Wendy Hartig-Merkel and Anna Nordle Gilliver translated the JPS into Swedish. Read the Swedish translation on the EMWA website: https://www.emwa.org/about-us/position-statements/joint-position-statement-on-predatory-publishing/swedish/.

EMWA web editorials

Have you checked out the web editorials page lately? Two new articles have been published recently. In modern society, everything we do needs to be quick, efficient and to-the-point. There is no doubt about it. Find out what Licia Genovese has to say on writing when you do not have time:

- https://www.emwa.org/about-us/emwa-news/web-editorial-archive/writing-when-you-don-t-have-time/

Online communication is changing. Discover how to use Hootsuite to manage your social media publishing by Diana Ribeiro:

- https://www.emwa.org/about-us/emwa-news/web-editorial-archive/how-to-use-hootsuite-to-manage-your-social-media-publishing/

We hope that you enjoy these short opinion pieces.

Changes to EMWA Freelance Directory

We recently made several changes to the layout of the Freelance Directory page, and members should now optimise their profile to obtain maximum search hits from potential new clients: https://members.emwa.org/EMWA/Freelance_Resources/Freelance%20Directory/Freelance_Directory.aspx

We encourage all freelance members to check whether their profiles accurately reflect experience and areas of medical writing in which they specialise. You should also add or update a profile photograph.

The new features include drop-down menus for the following categories:

- Language (select fluent language/s)
- Field of Expertise (Medical; Pharma; Scientific; Veterinary)
- Area of Expertise (Education Technology; Medical Communications; Medical Education; Regulatory)
- Specialist Services (Editing; Medical Devices; Medical Translation; Medical Writing; Proofreading; QA)
- Domain (select area(s) of medicine you work in)

Professional indemnity insurance discount for EMWA members

Did you know that EMWA members can get a 20% discount on their professional indemnity insurance?

Established in 1992, PIA Commercial works closely with its clients to provide a tailored range of specialist insurance products for both individuals and businesses. Please contact PIA Commercial at info@PIAcommercial.com for any queries or to receive a personalised quote.

Alternatively, visit their brand new updated website at *www.piacommercial.com* to view the extensive range of individualised insurance plans for businesses and individuals in the life science, biotechnology and healthcare industries. Keep up to date with their business news and industry insights by following them on LinkedIn.

EMWA Ambassador's Programme

Alison Rapley presented a workshop and a promotional talk for EMWA at the City University of London on March 12, just a week before the coronavirus lockdown. Altogether 20 students attended and gave very positive feedback.

EMWA's newest special interest group: Sustainability SIG

By the EMWA Sustainability Special Interest Group

In 2015, the United Nations set 17 Sustainable Developments Goals (SDGs, Figure 1) as a "universal call to action to end poverty, protect the planet, and ensure that all people enjoy peace and prosperity by 2030".[1] Central to these SDGs is planetary health, that body of research that looks at the complex interactions between "human-caused disruptions of Earth's natural systems and the resulting impacts on human health".[2]

As a professional organisation of medical communicators and healthcare professionals, we feel that EMWA should take a more active role in supporting these SDGs. Hence, we decided to ask for the EMWA Executive Committee (EC)'s blessing to set up a Special Interest Group (SIG) on Sustainability (SUS-SIG for short). We received approval on May 6, 2020.

Sustainability and health

There is a clear body of evidence indicating that global warming and the consequent climate change are drastically impacting human health. According to the World Health Organization, "between 2030 and 2050, climate change is expected to cause approximately 250,000 additional deaths per year, from malnutrition, malaria, diarrhoea and heat stress."[3] We are already experiencing unprecedented epidemics and natural disasters that directly and indirectly affect the health of whole human populations.

Governments and non-governmental organisations are taking steps in implementing policies to address these challenges but industries also need to do their part.

Many of us are employed within the pharmaceutical industry, a sector that is known for its very high carbon footprint[4] and large volume industrial waste.[5] Healthcare professionals, researchers, and scientists are calling for advocacy and action.[2] The World Medical Association, author of the Declaration of Helsinki, "is calling on all its members and on the global health community to adopt an environmentally responsible approach to their activities… This includes making health practice environmentally responsible and greening medical associations."[6]

On Earth Day last year, the Clinicians for Planetary Health Working Group issued a global call to action to prioritise planetary health.[2] The EAT-Lancet Commission on Food, Planet, Health released its first report based on a full scientific review of healthy diets and sustainable food systems to answer the question: Can we feed a future population of 10 billion people a healthy diet within planetary boundaries?[7]

Several professional associations and non-profit groups have also expressed their commitment to go green, such as PHUSE,[8] Green Nephrology for sustainable kidney care,[9] and the Korea Society for Green Hospitals.[6] It is also encouraging to see that more and more pharmaceutical companies are going beyond patient centricity towards planet centricity.[10] Through the SUS-SIG, we would like to mobilise the EMWA membership to support the SDGs professionally and privately.

Objectives of SUS-SIG

The objectives of this SIG are to:

- Promote and encourage action towards the 17 SDGs.
- Provide a forum for medical writers and communicators to discuss and share information in the area of sustainability.
- Support EMWA's commitment to reduce the carbon footprint of the medical writing and communication profession and the healthcare industry.
- Exchange tricks and tips on how to be envi-

ronmentally mindful healthcare professionals and individuals.

Potential activities

To achieve these goals, we have several activities lined up:

- Work with the EC and the EMWA Head Office to reduce the carbon footprint of EMWA conferences.
- Engage EMWA sponsors and service providers in conversation to support EMWA commitments to sustainability.
- Promote presentations, webinars, workshops, and other events related to sustainability.
- Contribute and solicit articles on sustainability for *Medical Writing*

But we need more ideas! This is a call to the EMWA membership for your support, ideas, and input. If interested, please contact the SIG Founders detailed below.

The SIG founders

Maria Carolina Rojido
carolinarojido@gmail.com
Carola Krause
Carola.Krause@codex-biomed.com
Raquel Billiones
medical.writing@billiones.biz

References

1. United Nations Development Program. Available from: https://www.undp.org/content/undp/en/home/sustainable-development-goals.html
2. Veidis EM, Myers SS, Almada AA, Golden CD; Clinicians for Planetary Health Working Group. A call for clinicians to act on planetary health. Lancet. 2019;393:2021.
3. World Health Organization. Climate Change. Available from: https://www.who.int/health-topics/climate-change#tab=tab_1
4. Belkhir L, Elmeligi A. Carbon footprint of the global pharmaceutical industry and relative impact of its major players. J Clean Prod. 2019;214:185–94.
5. Byrne M. How do you solve a problem like pharma waste? Gold EMC. February 2020. Issue 08:12-13. Available from: https://www.emg-gold.com/post/how-do-you-solve-a-problem-like-pharma-waste
6. World Medical Association. Green Health: Reducing the Environmental Impact of Healthcare. Available from: https://www.wma.net/what-we-do/public-health/green-health/.
7. EAT-Lancet Commission Summary Report 2019. Available from: https://eatforum.org/eat-lancet-commission/eat-lancet-commission-summary-report/
8. PHUSE Go Green. Available from: https://www.phuse.eu/phuse-go-green
9. Green Nephrology. Available from: https://networks.sustainablehealthcare.org.uk/network/green-nephrology.
10. Turner K. Gravitating towards planet centricity. Gold EMC. February 2020. Issue 08:9-11. Available from: https://www.emg-gold.com/post/gravitating-towards-planet-centricity

Resources

Below are some websites and links that provide reliable information and data on sustainability:

- UN Sustainable Development Goals: https://www.un.org/ sustainabledevelopment/ sustainable-development-goals/
- The One UN Climate Change Learning Partnership: https://www.uncclearn.org/
- WHO: https://www.who.int/ sustainable-development/en/
- Food and Agriculture Organization of the United Nations (FAO): http://www.fao.org/nutrition/ education/food-dietary-guidelines/background/ sustainable-dietary-guidelines/en/
- The Lancet Planetary Health: https://www.thelancet.com/journals/ lanplh/home

Figure 1. The 17 United Nations Sustainable Development Goals

Viewpoint

Rush to publication – *What do we have to lose?*

Just as the research and development of new drugs requires careful, often painstaking, adherence to empirical processes, the peer-review process and, indeed, the manuscript preparation process, are likewise laborious and time-consuming. The benefits of these are obvious and important, given how critical peer review serves as a "gatekeeper" for the disclosure of new scientific and medical information. However, we must consider weighing the potential value of rapid publication against the potential harm of inadequate vetting (both internal and external) of the final product.

In this time of global great peril where countless lives are held in the balance, are we willing to lower the threshold of scientific integrity for the sake of accelerating the availability of speculative medicinal products? This raises the topic of "situational ethics". Do desperate times require desperate measures? Added to the traditional dynamic tension between determining what is best for an individual vs. what is best for society at large, is the imminent threat of global pandemic for which there are few, if any, effective measures. Is "No Science" worse than "Bad Science"?

There is no question that bad science does not deserve a forum. However, good science needs to be heard even if some people will twist its meaning. Hopefully, scientists desire the safest and most effective treatment or vaccine and the most reliable diagnostic possible, but these cannot be refined if researchers ignore inconvenient data. Moreover, scientists will earn a lot more public trust, and overcome a lot more unfounded fear, if they choose transparency over censorship.

However, without an appropriate level of pre-publication vetting, how does one determine whether the article is based on good science? Do we have to wait until a more rigorous assessment after the genie is out of the bottle? I would argue that, at that point, the damage is done and no amount of retroactive "tagging" will have much effect. In a rush to "publish" studies that have not undergone traditional levels of scrutiny, unnecessary harm could easily result. Once "the toothpaste is out of the tube", it cannot easily (if, at all) be stuffed back in. Thus, in an online era, the misinformation is free to be circulated, cited, and believed *ad infinitum*, regardless of whether it is ultimately debunked and retracted. It should be noted that, at the time of this writing, Retraction Watch reports that 15 COVID-19 articles have been retracted, two temporarily retracted, and one has generated an expression of concern.[1]

The "poster child" example of the dissemination of fraudulent research findings is *The Lancet*'s 1998 publication of Andrew Wakefield's article linking the MMR (Measles-Mumps-Rubella) vaccine to autism – which, it should be noted, wasn't retracted until 12 years post-publication – and that was in the pre-open access, on-line era. Anti-vaxxers have taken to treating any attempts to discredit the Wakefield data as part of a conspiracy among a cabal, including the pharmaceutical industry, Bill Gates, and the "Deep State", intent on reaping huge financial gain at the expense of innocent children. Refusal to vaccinate results in a degradation of one of the founding principles of immunology – that of herd immunity. For example, if 80% of a population is immune to a virus, four out of every five people who encounter someone with the disease won't get sick (and won't spread the disease any further). In this way, the spread of infectious diseases is kept under control. Depending on how contagious an infection is, usually 70% to 90% of a population needs immunity to achieve herd immunity.

Should researchers handle findings differently when there is a chance they might frighten the public? Perhaps small, inconclusive, worrying studies should not be published because they could do more harm than good. Dr Paul Offit, director of the Vaccine Education Center at the Children's Hospital of Philadelphia states: "Knowing that you're going to scare people, I think you have to have far more data."[2]

One could argue that even an inconclusive paper can be important, as it can spur the larger, more definitive studies that are needed. It should be "put out there for the scientific community, to look at it, see it, know about it, refine study design and go and look again," says Gregory Poland, a Mayo Clinic vaccinologist and the Editor-in-Chief of *Vaccine*. It is crucial, though, for researchers to carefully explain such results in their papers to prevent misinterpretation. Even with appropriate disclaimers and cautions, however, nothing can prevent the "cherry-picking" of data to support one's particular *cause célèbre*.

The New York Times recently published an essay[3] in which the author noted:

> *As scientists race to understand the coronavirus, the process of designing experiments, collecting data and submitting studies to journals for expert review is being compressed drastically. What typically takes many months is happening in weeks, even as some journals are receiving double their normal number of submissions.*

The author brings into high relief how we should view the role of the medical/scientific journal:

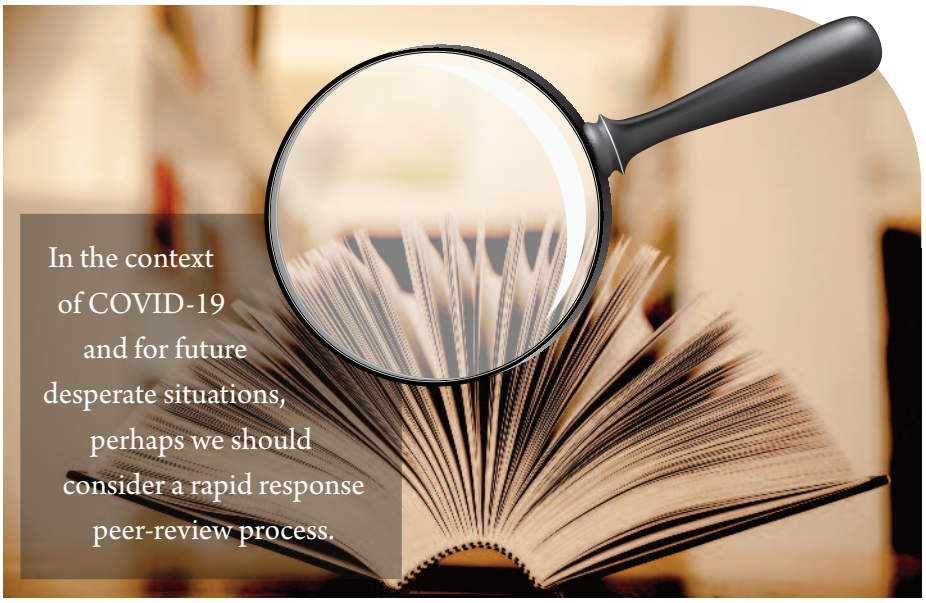
> *Should it be an arbiter of facts or a generator of new ideas? A keeper of the historical record or a predictor of the future? A private channel for scientists to communicate with one another or a megaphone with which they can reach the public? Or all of the above?*

In a world in which what is published today may strongly influence the practice of medicine, governmental policy decisions, and individual choices about social behaviour (e.g., mask-wearing, social distancing, resuming "normal" activities), releasing information that may be flawed, disingenuous, fraudulent, or politically influenced can have grave consequences. This is particularly true in an era fraught with conspiracy theorists who command huge audiences through on-line social media platforms.

Of course, there are even more egregious, and less-controlled pathways, of data release. Examples include the irresponsible (and unethical) Gilead teleconference, during which single-site data were shared and discussed among investigators, thereby undermining the principles and protections of Good Clinical Practice. I will not even delve into the promotion of completely unfounded claims from the podium by certain heads of state.

Improving the process

In the context of COVID-19 (and for future desperate situations), perhaps we should consider a "rapid response peer-review" process, comprising experts in applicable fields (virology, immunology, epidemiology etc.) who volunteer to drop everything at a moment's notice to give



at least a "cursory" peer review of any COVID-19 manuscript submitted to a journal. A 24-hour review deadline could be imposed and, there would at least be *some* assessment of the merits, pre-publication.

The post-publication peer-review process, adopted by F1000,[4] provides a pathway for peer-reviewed publication in as few as 14 days, with an in-house editorial team conducting a comprehensive set of prepublication checks to ensure that all policies and ethical guidelines are adhered to. Once the authors have finalised the manuscript, the article is published within a week, enabling immediate viewing and citation. However, a caveat is clearly communicated with a stamp noting that the article had not been peer-reviewed by the time of publication. The process next entails a phase of open peer review and user commenting (similar to Wikipedia). Expert reviewers are selected and invited, and their reports and names are published alongside the article, together with the authors' responses and comments from registered users. Authors are encouraged to publish revised versions of their articles. All versions of an article are linked and independently citable. Articles that pass peer review are indexed in external databases such as PubMed, Scopus, and Google Scholar. This process is sensible; however, it does not address a number of challenges associated with the urgency of the COVID-19 environment: insufficient speed of publication – most authors/institutions would be unwilling to delay publication by at least 14 days; and insufficient prestige – most authors/institutions would want to pre-publish/publish their findings in a prestigious journal. There may also be some issues regarding journal prior publication policies, potentially precluding publication in a journal if the manuscript was pre-published outside of that journal's auspices.

During health crises like COVID-19, the urgency of rapid publication may cause pre-publication in scientific journals, with post-publication peer review, to become the predominant pathway for medical researchers. However, we should be wary lest it become the norm under circumstances that may not warrant the relaxation of standard critical vetting processes. It is here that professional medical writers can serve as advisors and remind colleagues that there are well-established processes that should be followed.[5] These usually entail independent critical review, which will go a long way toward better ensuring the scientific quality and integrity of published research.[6]

Checking sources is also important: perhaps more credence can be given to information that comes from respected journals. But it is equally important to remember that even the best peer-reviewed advice is likely to change – and change again – particularly given the "black box" nature of this virus.

Ultimately, it is incumbent upon all of us who are intimately involved in the process of communicating science and medicine to caution against "first-blush" credibility. At the same time, we must not undermine the integrity of quality research findings, even if rapidly published. With some chagrin, I quote Ronald Reagan: "Trust but verify". Hopefully, in the final analysis, more good quality will prevail, and we will instil the place of value in a world of facts.

References

1. Retraction Watch [cited 2020 May 09]. Available from: https://www.RetractionWatch.com
2. Moyer MW. Anti-Vaccine Activists Have Taken Vaccine Science Hostage. The New York Times, 2018 Aug 4; Section SR, p. 7.
3. Tingley K. Corona virus is causing medical research to speed up. The New York Times Magazine, 2020 April 26; The Magazine, p. 16.
4. F1000 Research [cited 2020 May 09]. Available from: https://www.F10000research.com
5. EMWA/AMWA/ISMPP Joint position statement for professional medical writers [cited 2020 May 11] Available from: https://www.emwa.org/about-us/position-statements/joint-position-statement-for-professional- medical-writers
6. Wooley KL, Lew RA, Stretton S, et al. Lack of involvement of medical writers and the pharmaceutical industry in publications retracted for misconduct: a systematic, controlled, retrospective study. Curr Opin Med Res. 2011:27(6) 1175–82.

Art Gertel
medscicom@rcn.com

Author information

Art Gertel is an independent consultant, focusing on strategic regulatory guidance and bioethics. He brings over 40 years of increasingly senior-level positions in the pharmaceutical industry as well as leadership roles in professional organisations. Art is involved in numerous collaborative efforts to achieve improvement of the research, development, review, and approval of new therapeutics, vaccines, and diagnostics. He is a Registered Agent to the US FDA. He is proud to be a Fellow of EMWA and is a past president and Fellow of AMWA.

The value of registry data in the clinical evaluation of medical devices

Kelly Goodwin Burri[1], **Adrian Spoerri**[2]

[1] Stryker GmbH, Selzach, Switzerland
[2] SwissRDL, Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland

Correspondence to:

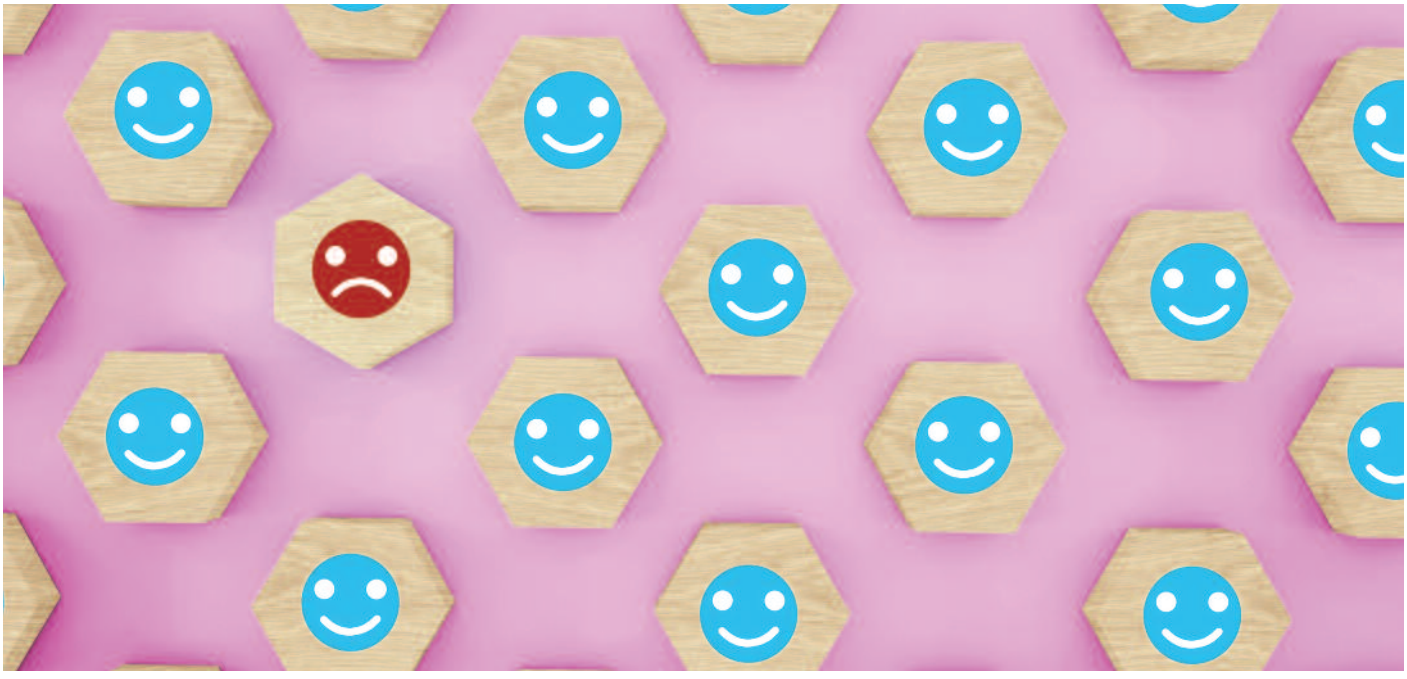
Kelly Goodwin Burri
Stryker GmbH
Bohnackerweg 1
2545 Selzach, Switzerland
kelly.goodwinburri@stryker.com

Abstract

Medical device manufacturers must continuously evaluate all clinical data available for their products marketed in Europe. With the European Medical Device Regulation 2017/745 coming into force in May 2021, manufacturers are required to assess available implant registry data as part of the clinical evaluation process. This new requirement will necessitate a closer collaboration between industry and registries to evaluate the safety and performance of high-risk devices. Medical writers should be aware of existing implant registries, understand what characteristics make a registry suitable to support regulatory requirements, and recognise both the value and the limitations of registries as a source of clinical evidence.

Patient registries provide a rich source of data on specific diseases, conditions, treatments, and exposures. Registry data are used to evaluate real-world treatments and outcomes, compare safety and effectiveness of treatments, monitor long-term safety, identify risk factors, and to assess quality in health care systems. Although randomised controlled trials are considered the gold standard for evaluating most medical treatments, it is not always feasible or ethical to carry them out on medical devices. Registries are increasingly seen as a supplement to data from randomised clinical trials, and in some cases, may be the only feasible approach to evaluate the long-term safety of some implantable devices.

The push to make better use of registry data for device surveillance increased as serious concerns about medical device safety came to light in 2012, specifically around the use of non-medical grade silicone in breast implants by Poly Implant Prothèse. This led the European Commission and EU countries to establish a joint action plan (Joint Plan for Immediate Actions under existing Medical Devices Legislation).[1] One of the five "immediate actions" was to support the development of implant registries that could identify safety issues and allow for long-term monitoring of safety and performance. Further implant safety issues making headlines in recent years, such as those related to metal-on-metal hip implants and vaginal mesh, only increased the pressure for heightened oversight and surveillance of medical devices, including calls for compulsory registration of all implantable devices.[2]



The European Medical Device Regulation (MDR) 2017/745 is the first regulation to include a specific requirement to evaluate registry data.[3] Article 108 of the MDR encourages the establishment of registries and registry networks based on common principles that enable the collection of comparable data on the long-term safety and performance of devices. It also suggests that registries contribute to traceability of implantable devices. In addition, the MDR requires both manufacturers and notified bodies to consider registry data as part of their obligations. Annex III (1.1 (a)) lays out the requirements to consider relevant databases and registries as part of post-market surveillance plans, while Annex VII (4.11(h)) requires an assessment of data from registries to be considered for re-certification by notified bodies.

The MDR requires both manufacturers and notified bodies to consider registry data as part of their obligations.

What is an implant registry?

The International Medical Device Regulator Forum defines a medical device registry as an "an organized system with a primary aim to increase the knowledge on medical devices contributing to improve the quality of patient care that continuously collects relevant data, evaluates meaningful outcomes and comprehensively covers the population defined by exposure to particular device(s) at a reasonably generalizable scale (e.g., international, national, regional, and health system)."[4] Registries use observational methods to collect standardised data on a population defined by a specific disease or condition (e.g., multiple sclerosis registry) or treatment with a specific product or procedure (e.g., arthroplasty registries). They may operate internationally, nationally, regionally, or at a single healthcare institution.[5] Many implant registries are led by professional medical societies or consortia. Implant registries collect more than just information about implants and include detailed information about patient characteristics and clinical outcomes.

Identifying implant registries

Keeping a complete and up-to-date overview of all operating implant registries is challenging. Generally well-established and long-standing national registries, such as the Swedish Knee Arthroplasty Register operating since 1975, the NJR (National Joint Registry) since 2002 in the UK, or SIRIS (Swiss National Joint Registry) since 2012, are easy to identify and will have the most valuable data in terms of quantity and quality. Smaller and newer registries may be identified through professional medical societies. Systematic reviews have been undertaken to map the implant registry landscape in Europe.[6,7] In 2013 researchers identified 101 implant registries in Europe and found that most are concentrated in the fields of cardiology (38 registries) and orthopaedics/arthroplasty (29 registries).[7]

A later review published in 2017 identified 24 hip and knee replacement registries in Europe.[6] Registries dedicated to other types of devices were less common: pacemakers and heart stents, breast implants, cochlear implants, insulin pumps, tubes, other stents, ophthalmological devices, brain stimulation/shunts, sacral neuromodulation, drug depots, and dental implants.[7] Another approach to identifying an appropriate registry is to search a registry of patient registries. The EMA inventory of registries maintained by the ENCePP (European Network of Centres for Pharmacoepidemiology and Pharmacovigilance)[8] and the cross-border PARENT (PAtient REgistries iNiTiative)[9] list of registries of Europe are two such initiatives that aim to increase transparency, avoid duplication, and promote collaboration among registries.

What kind of data can implant registries supply?

Registries can supply data at the level of the patient, implant, healthcare provider, and healthcare clinic, all factors that can influence clinical outcomes. Typical types of data collected include patient demographics (e.g., age, sex, comorbidities), procedure details (e.g., type of surgery, surgical approach, surgery duration), clinical data (e.g., indication, diagnosis, previous interventions), patient follow-up, adverse events, and implant details (e.g., UDI-DI [Unique Device Identification-Device Identifier] and device characteristics). However, more variables are not necessarily better, and the overall burden on data providers needs to be considered. Registries that collect fewer variables with an easy and quick procedure may better ensure the continued motivation of the data providers and a higher level of data quality. When assessing the appropriateness of an implant registry as a source of clinical evidence, the use of harmonised implant categorisation, patient-reported outcomes, and the ability of the registry to link to other data sources are especially important aspects to consider.

Implant data

Historically, registry data collection focussed on treatment procedures and outcomes in clinical files and patient records. With increasing use of implants, the focus shifted to device-related outcomes. To enable meaningful comparisons between similar implants, registries characterise each implant by collecting detailed information on product characteristics such as type, size, shape, material, coatings, or other important attributes. Several initiatives at the European level as well as globally have led to harmonised classification systems for orthopaedic implants, an important step to enable comparisons between registries. The implant library developed by the EPRD (German Arthroplasty Registries) and adopted by the NJR,[10,11] and the ISAR (International Society of Arthroplasty Registries) International Prosthesis Library[12] are examples of such efforts. In the example of arthroplasty, the use of a standard implant classification to analyse revision rates and implant survival is a prerequisite for a registry to serve as an early warning system for implant failures.

Patient-reported outcomes

The patient's subjective evaluation of healthcare outcomes using patient-reported outcome measures (PROMs) has gained recognition in value-based healthcare assessment. In quality assessment studies, patients are asked to complete a PROM questionnaire before a surgical intervention – for example about levels of pain, difficulties in daily activities, work-related limitations due to a health issue, and effects on social activities and family. After surgery, the same questionnaire completed by the patient at a specified follow-up time or multiple times (e.g., at 6 months and 1 year) is compared with the baseline measures to assess if the intervention was successful. Many registries are incorporating PROMs and recognise that these outcomes complement the clinical outcomes.

Enriched data through linkage

Data collection for registries adds to the administrative workload of health workers, and registries should be designed with only the

minimal information needed. Registries that are able to capture identifying patient information, where legal regulations and informed consents allow, enable future linkage to external datasets and reduce the burden on the registry data provider. For example, information about a patient's vital status (i.e., dead, alive, emigrated, or unknown) is crucial to calculate accurate revision rates in arthroplasty but is generally not available in registry data. Linkage of registry data with routinely collected administrative data, like mortality data, can overcome this limitation. Linkage to other types of datasets enrich the registry data and can facilitate analyses of important topics such as cost-effectiveness. Furthermore, electronic patient records or data of healthcare insurances are rich sources of information for quality assessment and research.

Suitability of implant registries for regulatory submissions

The availability of relevant data from a registry is just one aspect to consider when assessing the suitability of a registry to provide clinical evidence. The International Medical Device Regulator Forum Registry Working Group has defined 15 registry requirements, grouped into six elements, to assess the suitability of registry data for regulatory submissions (Table 1). The importance of each element is weighted differently depending on the intended use of the data. For example, the use of controlled vocabularies is recommended for post-market surveillance, while it is highly recommended for data intended to support an initial device approval or indication expansion.[13] Additional aspects that merit consideration are the completeness of data collection, transparent quality assurance processes, a clear policy for data access and sharing, and registry sustainability.[14]

An example – the Swiss National Implant Registry

SIRIS began registering hip and knee implants in September 2012 and is now the largest implant registry in Switzerland, with data collection supported by 186 healthcare institutions. Participation is compulsory for all hospitals and clinics performing knee and hip arthroplasties. The registry included 90%–92% of all hip and knee replacement procedures occurring in Switzerland, according to the most recent coverage estimates.[15] In the SIRIS 2019 annual report, implant types and brands have been compared for the first time using an implant library based on product catalogues from industry partners. In arthroplasty, the implant revision rate is the main outcome of interest. A revision procedure occurs when a patient's primary hip or knee primary implant is replaced by new components.

Figure 1 shows an example of the kind of analysis that can be performed with registry data with sufficiently detailed implant data collection. The funnel plot shows the 2-year revision rates for each participating health service unit by volume of performed operations. This analysis identifies clinics with revision rates outside of the expected variability by chance. A second type of analysis of specific product brands can identify "outlier" devices or device combinations that have a higher than expected revision rate than similar benchmark devices (Figure 2). However, because registries use observational study methods, many factors could contribute to an outlier status such as patient selection, case-mix, surgical technique, surgeon experience, and health service characteristics. The initial analysis shown in Figure 2 provides an alert that initiates a more in-depth analysis of the underlying cause of a poor outcome.[16]

Challenges with implant registry data

There are many challenges associated with registry data collection, and many are not limited specifically to implant registries. Case coverage, completeness, and data quality are relevant for all types of registries. High quality medical device data collection poses a unique challenge to implant registries as well as disease registries that attempt to collect implant data.

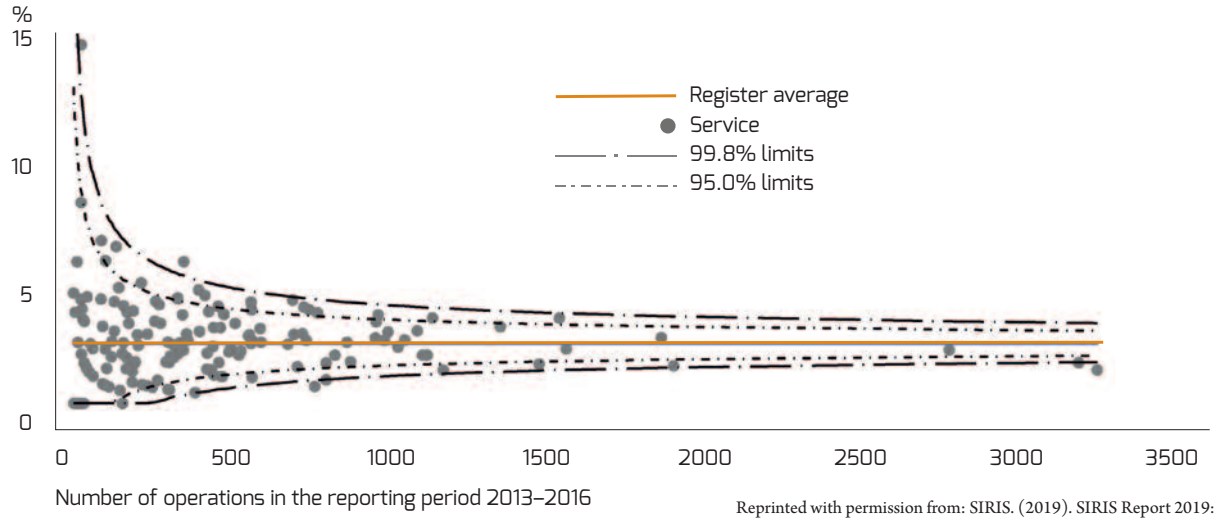
Coverage and completeness

To answer questions about quality of health care treatments or safety of medical devices using registry data, careful evaluation of potential sources of bias is paramount. High external validity, especially when compared with randomised clinical trials, is an important advantage of registries. Results from registry data are

Table 1. Registry elements affecting the suitability of clinical data for regulatory use

Element	Registry requirements
Governance	• Transparent governance structure and processes
Quality management system	• Legal requirements for data collection and handling are met • Information on patient data protection • Policy on access to data • Essential information available for verification (e.g., by competent authority, notified body)
Data gathering	• Relevant variables • Unambiguous device identification (e.g., UDI system) • Ability to link with other data sources • Use of controlled vocabularies • Use of harmonised minimum data model
Data storage	• Security protection against hacking, altering, deleting, or stealing data
Methodology/data analysis	• Conduct of analyses across different types of analysis frameworks • Data interpretation
Transparency/display/distribution	• Publicly available reports; report frequency and content • Publicly accessible website and web-reporting

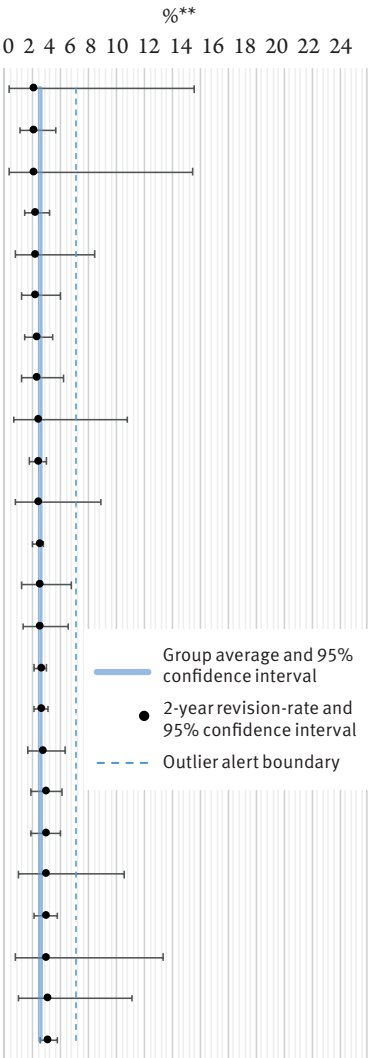
Adapted from the International Medical Device Regulator Forum Registry Working Group.[13]



Reprinted with permission from: SIRIS. (2019). SIRIS Report 2019: Annual Report of the Swiss National Joint Registry, Hip and Knee, 2012 – 2018. [15]

Figure 1. Example clinic analysis from SIRIS: Two-year revision rate of primary total hip arthroplasty by health care service

Combination	N revised	N at risk*
Exception + Betacup	1	50
Fitmore + RM pressfit vitamys	11	551
Harmony + April Ceramic	1	50
AMIStem + Versafitcup System	21	992
Alloclassic + Alloclassic	3	142
SL-plus + HI	10	468
Polarstem + Polarcup	19	882
SBG + R3	9	405
Corail + RM pressfit	2	88
Quadra + Versafitcup CC Trio	51	2278
Optimys + Allofit	3	134
Corail + Pinnacle	167	7124
Exception + Avantage	8	339
Quadra + Versafitcup System	10	416
AMIStem + Versafitcup CC Trio	145	5796
Avenir + Allofit	105	4164
SL-Plus + EP-fit	15	581
CLS + Allofit	25	895
CLS + Fitmore	28	1019
Corail + Allofit	3	107
Twinsys + RM pressfit vitamys	41	1485
Accolade + Trident	2	68
Exception + Exceed	3	99
Fitmore + Allofit	100	3301



* Number of patients with at least two years follow-up (i.e. primary prosthesis in 2012–2016).
** Rates ajusted for effects of mortality and emigration.

Figure 2. Example benchmark analysis from SIRIS: Two-year revision rates of uncemented stem-cup combinations used in primary total hip arthroplasty (2012–2018).

Reprinted with permission from: SIRIS. (2019). SIRIS Report 2019: Annual Report of the Swiss National Joint Registry, Hip and Knee, 2012 – 2018. [15]

generalisable if several conditions are met. A national registry needs to include all health services in the country delivering the treatment in focus. Underrepresentation of some areas or types of health services may introduce bias. Within a healthcare facility, all procedures meeting the inclusion criteria for the registry need to be recorded for full coverage. Excluding services or complex cases leads to bias in the analyses, interpretation of data, and generalisability. To calculate many outcome measures (e.g., the revision rate from arthroplasty registries), detailed knowledge about the registry coverage is vital. For example, if an implant revision surgery is performed in a clinic that does not record the operation in the registry, the revision rate will be underestimated. Another condition for unbiased analyses is the completeness of data. For the complete recording of implants, smart implant interfaces are needed. The type of operation (e.g., total hip arthroplasty) defines the expected type and number of implants and can be tracked during the scanning process. Warnings and error messages help to ensure that all expected implants for each case are captured.

Data quality

Several measures help to ensure and evaluate the quality of data and results:

1. use of reference data, sales figures, insurance data, or routinely collected administrative data to estimate the coverage of the registry,
2. precise definitions of inclusion and exclusion criteria for the registry, and
3. thoughtful design of electronic data capture forms, with precise definitions of variables, ranges of valid data, distinct categories of answers, mandatory and optional fields, and handling of potential missing data.

Measures for high registry data quality, coverage, completeness, and correctness can be implemented during different phases of the registry data capture process. Variable definitions, inclusion and exclusion criteria, and validation rules are defined before the data entry. During data entry, registry system rules provide warnings and errors, and first level support teams help with completing data entry forms. After data entry, automated monitoring routines and plausibility checks help detect potential errors or inconsistent data entries. Finally, for registries with sufficient funding, monitoring visits in the clinics and standardised audits verify the correctness of the data entered by comparing the source information in the clinical records with data captured in the registry.

Implant libraries

To access usable data for manufacturers to fulfil clinical evidence requirements, many registries do not have sufficiently detailed data collection to enable sophisticated analyses of specific implants. For example, some registries may collect data on general types of medical devices or implants used (e.g., plates, screws, external fixator) but not details that allow identification of a specific brand, model, or reference number. Another challenge is implementing a standardised categorisation of implants so that data may be compared across registries. Recent international congresses and meetings, for example the International Society of Arthroplasty Registries conferences, have advanced the discussion about standardisation and harmonisation of implant libraries. This led to agreements between the NJR in the UK and the German EPRD to harmonise their existing implant library definitions. Keeping these libraries up-to-date and accurate requires commitment from industry, with manufacturers needed to classify existing and newly marketed products according to a standard system with sufficient granularity that meaningful data analysis can happen.

In the future, the standardisation of implant libraries will reduce the administrative burden for manufacturers who provide implant catalogues with different categories and levels of granularity for different registries in many different countries. Ideally, implant registries may update their implant libraries using comprehensive implant data warehouses such as EUDAMED (European Database on Medical Devices) or other international databases. Unfortunately, local legal regulations leading to products sold in some but not other countries and challenges in standardisation processes hamper the development of international implant registries.

It is important to ensure that registries used to support regulatory requirements are well designed to produce valid data.

Conclusion

An increasing focus on the role and value of registries has led to steps to encourage better integration of registry data into regulatory decision making.[16,17] This effort requires the collaboration and input of all registry stakeholders, including patients, health care providers, professional societies, registry custodians, researchers, reimbursement bodies, public health and regulatory bodies, and the medical device industry. It is important to ensure that registries used to support regulatory requirements are well designed to produce valid data. The medical writer will play an important role in communicating clinical evidence on devices generated from registry data.

Disclaimers

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

Conflicts of interest

The authors declare no conflicts of interest.

References

1. European Commission. Medical devices: European Commission calls for immediate actions - tighten controls, increase surveillance, restore confidence. 2012 [cited 2020 Feb 29]. Available from: https://ec.europa.eu/commission/presscorner/detail/en/IP_12_119
2. Coombes R. Surgeons call for compulsory registers of all new medical devices. BMJ. 2018;363:k5010.
3. 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. [cited 2020 April 21]. Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1587480558555&uri=CELEX:32017R0746
4. IMDRF Patient Registries Working Group. Patient Registry: Essential Principles (IMDRF/Registry WG (PD1)/N33R1). 2015 [cited 2020 Feb 29]. Available from: http://www.imdrf.org/docs/imdrf/final/consultations/imdrf-cons-essential-principles-151124.pdf
5. Niederlander CS, Kriza C, Kolominsky-

Rabas P. Quality criteria for medical device registries: best practice approaches for improving patient safety – a systematic review of international experiences. Expert Rev Med Devices. 2017;14(1):49–64.
6. Lubbeke A, Silman AJ, Barea C, Prieto-Alhambra D, Carr AJ. Mapping existing hip and knee replacement registries in Europe. Health Policy. 2018;122(5):548-57.
7. Niederlander C, Wahlster P, Kriza C, Kolominsky-Rabas P. Registries of implantable medical devices in Europe. Health Policy. 2013;113(1-2):20–37.
8. ENCePP Resources Database. 2020 [cited 2020 Feb 29]. Available from: http://www.encepp.eu/encepp/resourcesDatabase.jsp
9. Patient registries of Europe. 2020 [cited 2020 Feb 29]. Available from: https://ec.europa.eu/health/sites/health/files/ehealth/docs/patient_registries_guidelines_en.pdf
10. Jansson V, Grimberg A, Melsheimer O, Perka C, Steinbruck A. Orthopaedic registries: the German experience. EFORT Open Rev. 2019;4(6):401–8.
11. Blomer W, Steinbruck A, Schroder C, et al. A new universal, standardized implant database for product identification: a unique tool for arthroplasty registries. Arch Orthop Trauma Surg. 2015;135(7):919–26.
12. Denissen GAW, van Steenbergen LN, Lollinga WT, Verdonschot NJJ, Schreurs BW, Nelissen R. Generic implant classification enables comparison across implant designs: the Dutch Arthroplasty Register implant library. EFORT Open Rev. 2019;4(6):344–50.
13. IMDRF Patient Registries Working Group. Tools for Assessing the Usability of Registries in Support of Regulatory Decision-Making (IMDRF/Registry WG/N46 FINAL:2018). 2018 [cited 2020 Feb 29]. Available from: http://www.imdrf.org/docs/imdrf/final/technical/imdrf-tech-180327-usability-tools-n46.pdf
14. McGettigan P, Alonso Olmo C, Plueschke K, et al., Patient Registries: An Underused Resource for Medicines Evaluation: Operational proposals for increasing the use of patient registries in regulatory assessments. Drug Safety. 2019;42(11):1343–51.
15. Annual Report of the Swiss National Joint Registry, Hip and Knee, 2012 – 2018. 2019 [cited 2020 Feb 29]. Available from: https://www.siris-implant.ch/de/News/Newsmeldung?newsid=14
16. Wilkinson J, Crosbie A. A UK medical devices regulator's perspective on registries. Biomed Tech (Berl). 2016;61(2):233–7.
17. Melvin T, Torre M. New medical device regulations: the regulator's view. EFORT Open Rev. 2019;4(6):351–6.

Author information

Kelly Goodwin Burri has 15 years of experience in bioengineering research, clinical development, epidemiology, and registry management. She currently works as a Senior Medical Writer for Stryker GmbH preparing clinical evaluation reports and related documentation for medical devices in the fields of orthopaedics and traumatology.

Dr Adrian Spoerri is head of SwissRDL, a centre of excellence for medical registries and data linkage at the University of Bern. He has long-standing experience in establishing and maintaining large cohorts and medical registries, providing reports and analyses for registry partners, and medical device reports for the industry.

Big data in clinical research: Present and future

Hyunjoo Kim[1], **Zuo Yen Lee**[2], **Chow Lih Yew**[3]

[1] Clinipace Clinical Research, Seoul, Republic of Korea
[2] Clinipace Clinical Research, Taipei, Taiwan (ROC)
[3] Clinipace Clinical Research, Singapore

Correspondence to:

Chow Lih Yew
Clinipace Clinical Research, Singapore
30 Cecil Street #19-08
Prudential Tower, Singapore 049712
+65 68275602
CYew@clinipace.com

Abstract

The clinical research landscape is gradually changing as we enter the era of big data. Big data sources are multiplying as existing sources collide to create expanded platforms that serve wider areas of expertise. Clinical study designs incorporating big data have started to appear and we expect this design phenomenon to grow. Big data offers unprecedented advantages in clinical research, but much remains to be done in assuring accessibility, validity, quality, and privacy protection. For these reasons, medical writers must understand big data, the strengths and the potential limitations of the data used, and should consider big data impact on study design, protocol, and clinical study report authoring. This article provides an overview of big data sources and provides insights on how big data utility could change the clinical-regulatory medical writing landscape.

The changing landscape of clinical research

The overall low generalisability of clinical trial results to routine clinical practice requires new approaches in clinical research.[1] Today, increasing data breadth and depth coupled with advancing data science offer new ways to assess a medicinal product across multiple data sources and at every step of the product's life cycle. We are entering the era of big data. EMA defines big data as "extremely large datasets which may be complex, multi-dimensional, unstructured and heterogeneous, which are accumulating rapidly and which may be analysed computationally to reveal patterns, trends, and associations".[2]

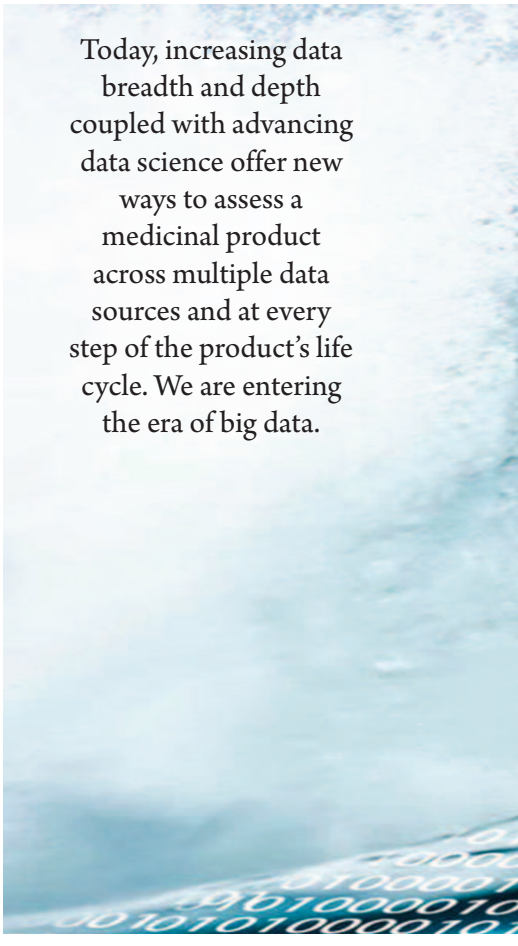
Myriad big data sources are now available, including those considered fit for regulatory decision-making. Table 1 lists example data sources – from the most traditional to relatively newer ones together with their main strengths and limitations.[3–11] This article discusses some of these data sources that are being actively applied in trials.

New ways to use patient registries

Patient registries are "organised systems that use observational methods to collect uniform data over time to evaluate specified outcomes for a population defined by a particular disease, condition, or exposure".[7,12,13] Patient registries could be a powerful tool in clinical studies as we see in VALIDATE-SWEDEHEART (clinicaltrial.gov number: NCT02311231), a prospective study that used the Swedish Coronary Angiography and Angioplasty Registry for both primary data (data collected for a specific, planned study, such as those of randomised clinical trials [RCT]) and secondary data (data already available for another purpose, such as insurance claims data) collection.[14,15] The study used the registry to assess and enrol potential subjects; collect their demographic and baseline data; and randomise subjects to treatment of percutaneous coronary intervention with either bivalirudin or heparin. After treatment, no study visit was required. All study data, including death, myocardial infarction, and major bleeding, were collected directly from the registry, via telephone calls and hospital records.[15,16] This study showcased the advantages of a registry-based RCT in which investigators could enrol many more subjects in a shorter time and the study gained both internal and external validity through the robust design of an RCT that utilised a data source (registry) with higher generalisability than a more traditional design would confer.[17]

Another advantage of working with registries is the accessibility to clinical data for rare diseases, and in which RCTs are often considered unfeasible.[18] Regulators recognise this; in one particular example, due to the low availability of previously untreated haemophilia A patients, the obligation to perform RCTs in these patients has been replaced for marketing authorisation applications of recombinant and human plasma-derived factor VIII products with a new requirement to monitor patients in a registry. The updated guideline also lists the core parameters to support homogeneous data collection across multiple registries – which should be taken into account at an early, pre-authorisation stage of study design. Registries may also be rich sources of secondary data from which suitable data could be extracted to serve as external controls, identify eligible patients, prevent duplicative data

Today, increasing data breadth and depth coupled with advancing data science offer new ways to assess a medicinal product across multiple data sources and at every step of the product's life cycle. We are entering the era of big data.



collection in clinical trials, and provide additional data for benefit/risk assessments.[19]

Using social media for pharmacovigilance

Data from social media are unique because they come directly from patients who have actively decided to share their information.[11] The use of social media in the field of pharmacovigilance and signal detection is not new. One example is the once-hyped Google Flu Trends, which was a web service introduced in 2008 providing estimates of influenza activity by analysing Google search queries. Google developed prediction models that could estimate influenza activity a couple of weeks ahead of the Centres for Disease Control and Prevention's periodic reports.[20] However, the service was terminated in 2015 after algorithmic glitches were detected.[21,22]

Despite Google's failure, numerous studies are testing new ways to utilise comments made on social media to identify potential adverse events; these studies suggest that social media is a promising tool for pharmacovigilance activities, but much work remains to determine its utility and validity.[23,24] Other areas of potential applications include data utilisation in effectiveness assessment, and as a communication tool to gather patient-centric data and to contact patients.[11]

Integrating mHealth in clinical studies

WHO defines mobile health, or mHealth, as the practice of medicine and public health supported by mobile devices for collecting data through symptom monitoring applications, implantable diagnostics, and wearable motion detectors.[25]

The 12-week exploratory Lilly Exploratory Digital Assessment Trial sponsored by Eli Lilly and Apple Inc. was conducted to explore how well mHealth data could discern those with mild cognitive impairment and early Alzheimer's disease from those free of these conditions. The model, applied to the data captured through distributed mobile phones and wearable devices, was able to discern patients from non-patients, suggesting that mild cognitive impairment could be detected in advance.[26] mHealth are in increasing use in clinical studies, acting as data sources for various real-time biometrics and other patient-reported outcomes. Although these novel modelling tools hold tremendous potential, they should be further assessed to ensure that they are "reliable, validated, reproducible, and predictable" to be used for the purpose of regulatory decision-making.[8]

Data collection through mobile devices will likely become more common in clinical studies following the release of the FDA's MyStudies App in 2018 – a digital platform used for multi-site or multi-database studies to collect primary data directly from patients' mobile devices. The application will be linked to an individual's electronic medical record (EMR) and enhanced with additional functions such as e-consent, eligibility test, survey delivery, notifications, and data validation.[27,28] It holds great potential for

Table 1. Sources of big data and their strengths and limitations

Data source	Strengths	Limitations
Clinical trial data (both interventional and non-interventional trials)[3]	• Well-structured data • High internal quality (integrity/veracity) • Non-selective sharing • Publicly accessible trial documents (e.g., protocol, statistical analysis plan)	• Data format and variable definitions across different trials are not standardised
Spontaneous adverse drug reports[4]	• System has a legal/regulatory framework • Data consistency at a global level • Competent in detecting new risks of medicines • Multi-dimensional data; various sources and safety concerns (e.g., medication errors, quality defects, cases of abuse/misuse, occupational exposure)	• Under/over-reporting • Risk of bias (the safety concern may be the result of increased media attention)
Drug consumption data[5]	• Cover large populations	• Lack individual patient data
Administrative claims data[5,6]	• Data consistency from standardised coding • Longitudinal record; in EU and in countries with public healthcare service, follow-up period is longer; representative for the source population at a national level • Provide linkage to data sources • High quality/complete drug exposure data • Data on individual's location available for geocoding	• Heterogeneous data in format, variables, quality, and completeness • Misclassification of diagnosis/exposure/outcome • Data might not be current • May lack data on secondary care • Lack of clinical details • Data protection legislation may prevent linkage between different health care providers • Lack of lifestyle/socio-economic factors; lack of control for confounding factors • Lack over-the-counter drug data
Electronic medical records (EMR)[5,6]	• Diverse clinical data; can complement claims data • Longitudinal in nature • Higher validity of diagnosis than claims data from routine use • Provide linkage to data sources	• Heterogeneous data in format, variables, quality, and completeness • Patient privacy concerns • May contain only one type of care setting (primary or secondary) • Lack of lifestyle/socio-economic factors; lack of control for confounding factors • Lack over-the-counter drug data
Patient registry [5–8]	• Data consistency • Established, large registry programmes • Able to observe the course of disease and effects of new treatments	• Limited to specific procedures, diseases, or settings • Data might not be current • Discrepancy between collected data and data requested by the regulatory authority • Inconsistent data and varying quality across registries • May need source data verification
Biomarkers (including any "omics" data)[8–10]	• Precision medicine • Identification of unique molecular markers of disease/responsiveness to medications	• High genetic variation • False positives/negatives • Further validation needed to associate biomarker data to patient outcomes • Lack of publicly accessible, clinically meaningful information from the genomic database • Lack of data standardisation • Patient privacy concerns, especially for patients with rare diseases • Heterogeneous data

Data source	Strengths	Limitations
Medical imaging[3,6]	• General data consistency • Widely used in clinical trials; unexplored potential in various therapeutic areas	• Lack of accessibility • Ethical issues related to data sharing • Challenges on analysing/integrating imaging data with other data sources
Social media[6,11]	• Wide reach of the internet • Various types of data • Result of active sharing from patients	• Heterogeneous data • Lack of specificity in general social media; data prone to bias • Limited follow up; difficult to verify/validate • Lacks consideration of the characteristics of the patients included • Lacks Good Clinical Practice adherence • Lacks validity and reliability • Patient privacy concerns
Mobile health (mHealth) and wearable devices[6,8,11]	• Collected biometrics data may allow control for confounding factors • Patient-centric data • Continuous data from real life (vs. episodic data restricted to healthcare setting) • Data readily available for research purposes; platforms support central data management, analysis, and reporting and can often be directly linked to an electronic case report form • Devices can monitor parameters to calculate/monitor drug dose	• Further validation needed to discern clinically important 'signals'; unknown sensitivity of the collected data • Precision does not necessarily mean accuracy • Output is highly variable across different types of device • Output depends on the level of user interaction • Lack of familiarity with interpretation of the data • Potential challenges in timing of surveys in relation to other healthcare data • Patient privacy and security concerns, e.g., hacking

pragmatic trials – which are evidential for the use of a clinical practice intervention and may, therefore, guide policy-making, and observational trials – trials without an intervention.[29]

In the wake of the recent coronavirus disease 2019 (COVID-19) pandemic, many mHealth initiatives have been developed for gathering information to help manage the outbreak. For example, Scripps Research Translational Institute launched the DETECT study in March 2020 to collect health data through wearables like smartwatches and activity trackers for a public health surveillance programme for early detection of viral diseases; at the same time, Stanford Medicine also initiated the COVID-19 Wearable Study that serves the same purpose.[30–33] In April 2020, the two platforms joined forces, together with Fitbit, to create a consortium which will aggregate data for knowledge sharing and drive wearables research.[32,34]

Larger data platforms

Existing data are expanding, and are also being linked across various networks, creating larger data platforms. Sentinel is FDA's national safety surveillance system to monitor its regulated medical products. The system extracts electronic health records (EHRs) from various networks, mostly from health insurers.[35] Sentinel is now collaborating with over 40 other networks across three centres – Sentinel Operations Centre, Innovation Centre, and Community Building and Outreach Centre – to cover wider areas of scientific expertise, improve technologies translation, and encourage communication and collaboration.[35]

A Sentinel collaborator, Patient-Centered Outcomes Research Network (PCORnet®), is a partnership of over 10 networks. PCORnet® contains more EMR data with various types of individual patient data, including laboratory test results, vital signs, biospecimen data, genomic data, and patient satisfaction data.[36,37] ADAPTABLE (clinicaltrials.gov number: NCT02697916), a pragmatic clinical study that

compares the effectiveness of two doses of aspirin (81 mg and 325 mg) in approximately 20,000 patients, uses the existing EHRs and a web-based patient portal in PCORnet® to identify eligible patients, obtain consents, randomise, and follow up with patients.[38]

No such platform is available in the EU yet. However, recently the Heads of Medicines Agencies (HMA)/EMA Task Force on Big Data proposed their plans to establish an EU platform, namely Data Analysis and Real-World Interrogation Network (DARWIN), to access and analyse healthcare data from across the EU to inform regulatory decision-making. This initiative is one of the many efforts undertaken by the EMA to optimise the use of big data in medicines regulation.[39] During the recent COVID-19 pandemic, international regulators and experts from WHO and European Commission acknowledged the value of real-world data from COVID-19 observational studies and how these data could complement clinical trials in finding solutions to prevent and treat COVID-19. Public platforms, such as EU PAS Register and ClinicalTrials.gov, were identified as suitable platforms to share and exchange information about COVID-19 observational studies.[40,41]

On the horizon

In the future, it may be possible to create a complete, longitudinal record of an individual starting from the omics level. Collaborations between academia, companies, and regulatory authorities nationally and internationally culminated in the initiative of the Electronic Medical Records and Genomics (eMERGE) Network. Since 2007, the Network has brought together researchers in genomics, statistics, ethics, informatics, and clinical medicine areas with the goal to combine a biological materials repository with EMR systems for research at the genomic level.[9,42]

About 90% of medical data are in the form of images captured with increasingly higher quality and improved resolution. Much of these voluminous data are stored unanalysed.[43] To utilise these data, the UK Biobank Imaging Study aims to develop longitudinal records from volunteers consisting of their brain, heart, and body imaging data; biomarker and genetic analysis results; physical measurements; and self-reported health and lifestyle data. These records can also be linked to the individual's National Health Service records.[44,45] In April 2020, the UK Biobank announced that it would grant access to the health data of its 500,000 participants to researchers for health-related research. These data include results of COVID-19 tests, primary care data, hospital episodes, and intensive care data.[46]

The exponential advances in personal omics profiling, coupled with the increasing amount of high-frequency data using wearable devices, omics data, imaging data, as well as enlarging platforms and dynamic patient-centred interfaces are set to greatly affect how we conduct clinical research.

Guidelines for using big data in regulatory decision-making

The message from regulators is that we must embrace the use of big data. In January 2017, the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use Assembly (ICH) endorsed the ICH reflection paper entitled "ICH Reflection on "GCP Renovation": Modernisation of ICH E8 and Subsequent Renovation of ICH E6"[47] to address the increasing diversity of clinical trial designs and data sources being employed.

FDA has also published their final guidance on the "Use of Real-World Evidence to Support Regulatory Decision-Making for Medical Devices" (August 2017),[48] "Use of Electronic Health Record Data in Clinical Investigations" (July 2018),[49] and a draft guidance for "Submitting Documents Using Real-World Data and Real-World Evidence to FDA for Drugs and Biologics" (May 2019)[50] to guide and encourage the use of big data in the industry.

> As regulatory medical writers, we need to consider how best to leverage big data into our work outputs.

Using big data in medical writing

Using big data to improve study efficiency

As regulatory medical writers, we need to consider how best to leverage big data into our work outputs. Big data could be applied at any stage of study design, through to enrolment and data analysis.[51] Before study initiation, we could use existing big data, sourced from registries and EHRs to help identify an appropriate target population, define more targeted eligibility

criteria, determine if a sufficient number of subjects are likely to be available, and even obtain baseline data directly from the data sources. Thus, study efficiency could be greatly improved by reducing the cost of and time for recruitment, reducing patient attrition and minimising possible changes to the protocol down the line.[51,52]

Throughout the study, data collected from various digital sources (such as mobile applications and wearable devices) may be available faster than data collected by traditional methods, thereby allowing for prompt futility analyses in a study or benefit/risk assessment in a post-marketing surveillance, hence more rapid decision-making. Big data platforms like registries also help track patients during study follow-up under their usual care routines, thus minimising patients being lost to follow-up and reducing missing data during a study.[51]

Using big data in study design

For some diseases where patient enrolment may be problematic (e.g., rare diseases) or randomising patients to the control group may be unethical (e.g., cancer), using an external control group can be considered. An external control group refers to subjects who are selected from an external source, e.g., existing clinical trial data and EHRs. The biggest challenge of using an external control is bias control. FDA suggests the use of external controls only under certain conditions, e.g., when we expect distinct treatment effects between the test and external control groups. External control should be selected from data sources that are most appropriate to the study purpose and should align, as much as possible, with the study eligibility criteria to minimise potential confounding and selection biases.[53] Another important consideration is the availability of similar endpoint assessments between the test group and the external control group to allow comparison between them. In this case, external control groups derived from existing clinical studies with similar purposes may be more applicable than those from EHRs or registries.[54]

Heterogeneity in the data is the intrinsic underlying issue in most data sources and this aspect should be thought through in the study design and statistical analysis, in consultation with biostatistical colleagues – our natural partners in analysis and reporting. When selecting which databases to use, accessibility, storage, and quality of the data are paramount considerations as they ensure reliability and validity of the data. We must be mindful while extracting data that they may contain missing information that could bias our interpretation of the data. For example, missing data does not mean the absence of an event; the absence of smoking status in the medical record may not mean the patient is not a smoker.

Heterogeneity in the data is the intrinsic underlying issue in most data sources and this aspect should be thought through in the study design and statistical analysis, in consultation with biostatistical colleagues – our natural partners in analysis and reporting.

Designing a study using big data requires rather different elements and methods from that of traditional RCTs. Existing guidelines such as the European Network of Centres for Pharmacoepidemiology and Pharmacovigilance "Guide on Methodological Standards in Pharmacoepidemiology",[14] its protocol checklist,[55] and the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) checklist[56] provide some starting guidance for medical writers on important elements that should be considered, such as procedures for selecting target population, defining covariates, methods to address each type of bias, and related statistical analyses.

It is important to remember that study design is always a collaborative endeavour with colleagues in other functional areas such as biostatistics and medical affairs. As medical writers, we can and should be influencers. We can raise awareness of the potential of big data in study design to ensure that all stakeholders consider its practical utility.

Permission from a subject to use his or her personal health data, in the form of informed consent or authorisation with pre-defined purposes, is required before data collection.

Permission for secondary use of personal data

Permission from a subject to use his or her personal health data, in the form of informed consent or authorisation with pre-defined purposes, is required before data collection. Big data analytics seek patterns and associations from big datasets that are often generated by pooling or linking data from various studies and databases. Therefore, secondary use, i.e. use of existing data collected for other purposes, is more common for big data analytics.

Personal data that will be collected, processed, or stored within the EU need to comply with the General Data Protection Regulation (GDPR). Under the GDPR, new consent is not required for the processing and secondary use of personal data for scientific research purposes provided specific adequate safeguards and conditions are adhered to, such as pseudonymisation.[57,58] GDPR also acknowledges that it is "often not possible to fully identify the purpose of personal data processing for scientific research purposes at the time of data collection" and allows subjects to consent to a more general purpose. Nonetheless, in addition to the specific consent, GDPR requires that a separate consent with the general areas of secondary research be specified and options to "consent only to certain areas of research or parts of the research projects" be provided before data collection.[59,60] Of note, the use of de-identified personal data does not fall within the scope of the GDPR.[61]

In the US, the "Revised Common Rule" that took effect in January 21, 2019, accelerates the secondary use of data through the introduction of Broad Consent. Broad Consent allows subjects to consent to unspecified future research that may store, maintain, or use their identifiable private information or identifiable biospecimens for secondary research before data collection. Important information, such as the types of research that may be conducted, information that may be used, the institutions that may reuse the information, and the time frame of the consent must be included in the Broad Consent.[62,63]

Under the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule, "covered entities", including health plans, health care clearinghouses, and health care providers, should obtain an individual's written authorisation for any use of protected health information (PHI) for secondary research.[64,65] Core elements, such as the purpose of the use, the specific information to be used, the persons who can use the PHI, and the time frame of the authorisation must be included in the

authorisation.[65] An Institutional Review Board or privacy board waiver of authorisation is required to use PHI for research purposes if individual authorisation is not available.[66] Currently, there are no restrictions (i.e. neither consent nor HIPAA authorisation is required) on the use of de-identified health information.[67,68]

Big data is expected to offer unprecedented advantages in every step of clinical research by providing alternative study design, improving study efficiency, and accelerating regulatory decision-making.

Conclusion

Big data is expected to offer unprecedented advantages in every step of clinical research by providing alternative study design, improving study efficiency, and accelerating regulatory decision-making. At the same time, they also pose new challenges, especially in ensuring data quality and privacy protection. An enormous amount of health data has become available during the recent COVID-19 pandemic, and we have directly experienced how researchers and regulators across the world use big data in the fight against COVID-19. Undoubtedly, we as medical writers should start honing the necessary skills and competencies to better prepare ourselves as we embrace the era of big data.

Acknowledgements

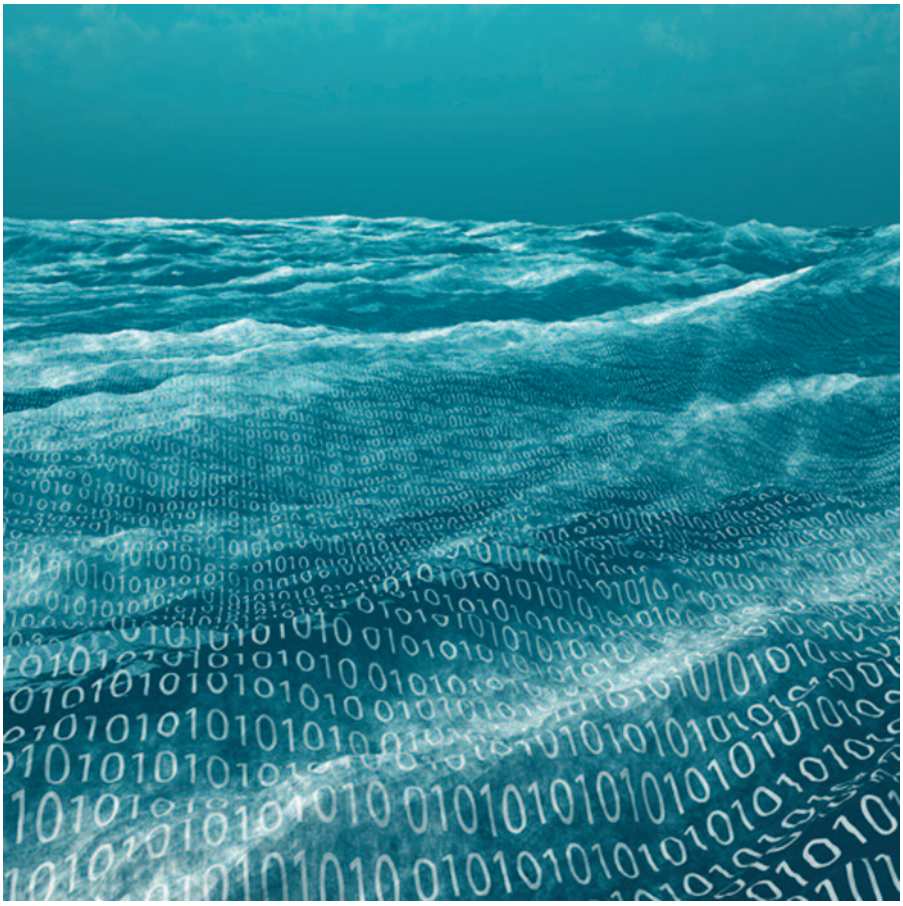
The authors would like to thank Samina Hamilton for her review and input in the draft manuscript.

Conflicts of interest

The authors declare no conflicts of interest.

References

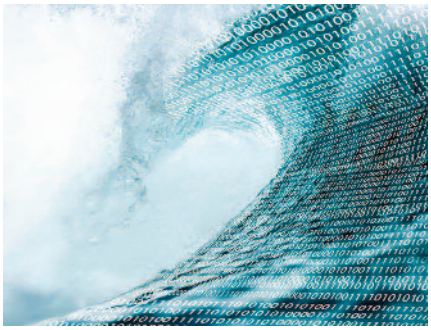
1. Bonell C, Oakley A, Hargreaves J, Strange V, Rees R. Assessment of generalisability in trials of health interventions: suggested framework and systematic review. BMJ. 2006;333(7563):346–9.
2. European Medicines Agency. Big data. 1995-2020 [cited 2020 Feb 26]. Available from: https://www.ema.europa.eu/en/about-us/how-we-work/big-data
3. Kiviniemi V, Rosso A, Szabone ZC, Nyeland M, Fuglerud P. Clinical Trial and Imaging Subgroup report [Internet]. [cited 2020 Feb 26]. Available from: https://www.ema.europa.eu/en/documents/report/clinical-trial-imaging-subgroup-report_en.pdf
4. García CH, Pinheiro L, Maciá MÁ, et al. Spontaneous Adverse Drug Reactions Subgroup report [Internet]. [cited 2020 Feb 26]. Available from: https://www.ema.europa.eu/en/documents/report/spontaneous-adverse-drug-reactions-subgroup-report_en.pdf
5. Rosso A, Pacurariu A, Cave A, et al. Observational data (Real World Data) Subgroup report [Internet]. [cited 2020 Feb 26]. Available from https://www.ema.europa.eu/en/documents/report/observational-data-real-world-data-subgroup-report_en.pdf
6. Rumsfeld JS, Joynt KE, Maddox TM. Big data analytics to improve cardiovascular care: promise and challenges. Nat Rev Cardiol. 2016;13(6):350–9.
7. Gliklich RE, Leavy MB, Dreyer NA, editors. Registries for Evaluating Patient Outcomes: A User's Guide. (Prepared by the Outcome DEcIDE Center [Outcome Sciences, Inc. dba Outcome] under Contract No. HHSA29020050035I TO1.) AHRQ Publication No. 07-EHC001-1 [Internet]. Rockville, MD: Agency for Healthcare Research and Quality; 2007 [cited 2020 Feb 26]. Available from: https://effectivehealthcare.ahrq.gov/sites/default/files/pdf/registries-guide_research.pdf
8. emainfo. Welcome, and update on work of the HMA/EMA Joint Big Data Task Force on Big Data [video file]. 2018 [cited 2020 Feb 26]. Available from: https://www.youtube.com/watch?list=PL7K5dNgKnawbkBU1mXLDWAnfkFOzSeDRo&v=mUDdJH0lf2I&feature=emb_logo
9. Meulendijks D, Deforce D, Ovelgönne H, König R, Pasmooij M. Genomics Genetics, Transcriptomics and Epigenetics Subgroup report [Internet]. [cited 2020 Feb 26]. Available from https://www.ema.europa.eu/en/ documents/report/genomics-genetics-transcriptomics-epigenetics-subgroup-report_en.pdf
10. König R, Cave A, Goldammer M, Meulendijks D. Bioanalytical Omics Subgroup report [Internet]. [cited 2020 Feb 26]. Available from: https://www.ema.europa.eu/en/documents/report/bioanalytical-omics-subgroup-report_en.pdf
11. Donegan K, Ovelgonne H, Flores G, Fuglerud P, Georgescu A. Social Media and M-Health Data Subgroup report [Internet]. [cited 2020 Feb 26]. Available from: https://www.ema.europa.eu/en/documents/report/social-media-m-health-data-subgroup-report_en.pdf
12. Zaletel M, Kralj M, editors. Methodological Guidelines and Recommendations for Efficient and Rational Governance of Patient Registries [Internet]. Ljubljana: National Institute of Public Health; 2015 [cited 2020 Feb 26]. Available from: https://ec.europa.eu/health/sites/health/files/ehealth/docs/patient_registries_guidelines_en.pdf
13. Polygenis D, Frame D, Blanchette C, editors. ISPOR Taxonomy of Patient Registries: Classification, Characteristics and Terms. Lawrenceville, NJ: ISPOR; 2013.
14. The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance. Guide on Methodological Standards in Pharmacoepidemiology (Revision 7). Reference No.: EMA/95098/2010 [Internet]. [cited 2020 Feb 26]. Available from: http://www.encepp.eu/standards_and_guidances/documents/ENCePPGuideonMethStandardsinPE_Rev7.pdf
15. ClinicalTrials.gov. Bivalirudin vs Heparin in NSTEMI and STEMI in Patients on Modern Antiplatelet Therapy in SWEDEHEART (VALIDATE). 2017 [cited 2020 Feb 26]. Available from: https://clinicaltrials.gov/ct2/show/NCT02311231
16. Erlinge D, Omerovic E, Fröbert O, et al. Bivalirudin versus Heparin Monotherapy in Myocardial Infarction. N Engl J Med. 2017;377(12):1132–42.
17. Lauer MS, D'Agostino RB Sr. The Randomized Registry Trial – The Next Disruptive Technology in Clinical Research?. N Engl J Med. 2013;369(17):1579–81.
18. McGettigan P, Alonso Olmo C, Plueschke



K, et al. Patient Registries: An Underused Resource for Medicines Evaluation: Operational proposals for increasing the use of patient registries in regulatory assessments. Drug Saf. 2019;42(11): 1343–51.
19. European Medicines Agency. Guideline on the clinical investigation of recombinant and human plasma-derived factor VIII products (Revision 2). Reference No.: EMA/CHMP/BPWP/144533/2009 [Internet]. [cited 2020 Feb 26]. Available from: https://www.ema.europa.eu/en/clinical-investigation-recombinant-human-plasma-derived-factor-viii-products
20. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature. 2009;457(7232):1012–4.
21. Cook S, Conrad C, Fowlkes AL, Mohebbi MH. Assessing Google flu trends performance in the United States during the 2009 influenza virus A (H1N1) pandemic. PLoS One. 2011;6(8):e23610.
22. Olson DR, Konty KJ, Paladini M, Viboud C, Simonsen L. Reassessing Google Flu Trends data for detection of seasonal and pandemic influenza: a comparative epidemiological study at three geographic scales. PLoS Comput Biol. 2013;9(10):e1003256.
23. Anderson LS, Bell HG, Gilbert M, et al. Using Social Listening Data to Monitor Misuse and Nonmedical Use of Bupropion: A Content Analysis. JMIR Public Health Surveill. 2017;3(1):e6.
24. Powell GE, Seifert HA, Reblin T, et al. Social Media Listening for Routine Post-Marketing Safety Surveillance. Drug Saf. 2016;39(5):443–54.
25. World Health Organization. mHealth: New horizons for health through mobile technologies (Global Observatory for eHealth series - Volume 3) [Internet]. 2011 [cited 2020 Feb 26]. Available from https://www.who.int/goe/publications/ehealth_series_vol3/en/
26. Chen R, Jankovic F, Marinsek N, et al. Developing Measures of Cognitive Impairment in the Real World from Consumer-Grade Multimodal Sensor Streams. In: The 25th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '19); 2019 August 4-8; Anchorage, AK, US. New York: ACM; 2019. p2145-2155. Available from https://evidation.com/wp-content/uploads/2019/08/developing-measures-of-cognitive-impairment-in-the-real-world-from-consumer-grade-multimodal-sensor-streams.pdf
27. U.S. Food & Drug Administration. FDA's MyStudies Application (App). 2019 [cited 2020 Feb 26]. Available from: https://www.fda.gov/drugs/science-and-research-drugs/fdas-mystudies-application-app
28. U.S. Food & Drug Admnistration. Mobile App Quick Overview for Research [Internet]. 2019 [cited 2020 Feb 26]. Available from: https://www.fda.gov/media/119834/download.
29. Ford I, Norrie J. Pragmatic Trials. N Engl J Med. 2016;375(5):454–63.
30. Scripps Research Translational Institute. Scripps Research invites public to join app-based DETECT study, leveraging wearable data to potentially flag onset of viral illnesses. [Internet]. 2020 [cited 2020 Apr 19]. Available from: https://www.scripps.edu/news-and-events/press-room/2020/20200325-detect-study-viral-illnesses.html
31. Scripps Research Translational Institute. DETECT Health Study. 2020 [cited 2020 Apr 19]. Available from: https://detectstudy.org/
32. Stanford Medicine. Stanford Medicine scientists hope to use data from wearable devices to predict illness, including COVID-19. [Internet] 2020 [cited 2020 Apr 19]. Available from https://med.stanford.edu/news/all-news/2020/04/wearable-devices-for-predicting-illness-.html
33. Stanford Medicine. Fight COVID-19 through the power of people. 2020 [cited 2020 Apr 19]. Available from: https://innovations.stanford.edu/#section-wearables
34. mHealth Intelligence. SRTI, Stanford Health Launch mHealth Consortium for COVID-19 Research. [Internet] 2020 [cited 2020 Apr 19]. Available from: https://mhealthintelligence.com/news/srti-stanford-health-launch-mhealth-consortium-for-covid-19-research
35. U.S. Food & Drug Admnistration. FDA's Sentinel Initiative. 2019 [cited 2020 Feb

26]. Available from https://www.fda.gov/safety/fdas-sentinel-initiative
36. PCORnet. PCORnet Common Data Model. 2019 [cited 2020 Feb 26]. Available from: https://pcornet.org/data-driven-common-model
37. PCORnet. Real World Data for Clinical Research: A PCORnet Workshop with the Pharmaceutical and Biologics Industry [Internet]. Washington, DC, US: PCORI; 2019 [cited 2020 Feb 26]. Available from: https://www.pcori.org/events/2015/real-world-data-clinical-research-pcornet-workshop-pharmaceutical-and-biologics-industry
38. Johnston A, Jones WS, Hernandez AF. The ADAPTABLE Trial and Aspirin Dosing in Secondary Prevention for Patients with Coronary Artery Disease. Curr Cardiol Rep. 2016;18(8):81.
39. European Medicines Agency. Ten recommendations to unlock the potential of big data for public health in the EU. 2020 [cited 2020 Feb 26]. Available from: https://www.ema.europa.eu/en/news/ten-recommendations-unlock-potential-big-data-public-health-eu
40. European Medicines Agency. Global regulators discuss observational studies of real world data for COVID-19 medicines. [Internet] 2020 [cited 2020 Apr 19]. Available from: https://www.ema.europa.eu/en/news/global-regulators-discuss-observational-studies-real-world-data-covid-19-medicines
41. European Medicines Agency. Meeting highlights from ICMRA global regulatory workshop on COVID-19 observational studies and real world data. [Internet] 2020 [cited 2020 Apr 19]. Available from: https://www.ema.europa.eu/en/news/meeting-highlights-icmra-global-regulatory-workshop-covid-19-observational-studies-real-world-data
42. National Human Genome Research Institute. Electronic Medical Records and Genomics (eMERGE) Network. 2020 [cited 2020 Feb 26]. Available from https://www.genome.gov/Funded-Programs-Projects/Electronic-Medical-Records-and-Genomics-Network-eMERGE
43. GE Healthcare. Beyond Imaging: the Paradox of AI and Medical Imaging Innovation. 2018 [cited 2020 Feb 26]. Available from: https://www.gehealthcare.com/article/beyond-imagingthe-paradox-of-ai-and-medical-imaging-innovation
44. Nature. UK Biobank data on 500,000 people paves way to precision medicine. 2018 [cited 2020 Feb 26]. Available from: https://www.nature.com/articles/d41586-018-06950-9
45. UK Biobank. About UK Biobank. 2019 [cited 2020 Feb 26]. Available from: http://www.ukbiobank.ac.uk/about-biobank-uk/
46. UK Biobank. UK Biobank makes infection and health data available to tackle COVID-19. [Internet] 2020 [cited 2020 Apr 19]. Available from: https://www.ukbiobank.ac.uk/2020/04/covid/
47. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use. ICH Reflection on "GCP Renovation": Modernization of ICH E8 and Subsequent Renovation of ICH E6. 2017 [cited 2020 Feb 26]. Available from: https://admin.ich.org/sites/default/files/2019-04/ICH_Reflection_paper_GCP_Renovation_Jan_2017_Final.pdf
48. U.S. Food & Drug Administration. Use of Real-World Evidence to Support Regulatory Decision-Making for Medical Devices. Reference No.: FDA-2016-D-2153 [Internet]. 2017 [cited 2020 Feb 26]. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/use-real-world-evidence-support-regulatory-decision-making-medical-devices
49. U.S. Food & Drug Admnistration. Use of Electronic Health Record Data in Clinical Investigations. Reference No.: FDA-2016-D-1224 [Internet]. 2018 [cited 2020 Feb 26]. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/use-electronic-health-record-data-clinical-investigations-guidance-industry
50. U.S. Food & Drug Administration. Submitting Documents Using Real-World Data and Real-World Evidence to FDA for Drugs and Biologics. Reference No.: 2019-09529 [Internet]. 2019 [cited 2020 Feb 26]. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/submitting-documents-using-real-world-data-and-real-world-evidence-fda-drugs-and-biologics-guidance
51. Mc Cord KA, Al-Shahi Salman R, Treweek S, et al. Routinely collected data for randomized trials: promises, barriers, and implications. Trials. 2018;19(1):29.
52. Gilder JR, Big Data in Healthcare: The next frontier in clinical research, study trial recruitment, and quality assessment. The Monitor. 2012;26:55–8.
53. U.S. Food & Drug Administration. E10 Choice of Control Group and Related Issues in Clinical Trials. Reference No.: FDA-2013-S-0610 [Internet]. 2001 [cited 2020 Feb 26]. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/e10-choice-control-group-and-related-issues-clinical-trials
54. Beckers F, Capra W, Cassidy A, et al. Characterizing the use of external controls for augmenting randomized control arms and confirming benefit. Proceedings of the Friends 2019 Annual Meeting Panel 1; 2019 Nov 12; Washington, DC, US. Washington, DC: Friends of Cancer Research. 2019 [cited 2020 Feb 26]. Available from https://www.focr.org/sites/default/files/pdf/Panel-1_External_Control_Arms2019AM.pdf
55. The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance. Checklist for Study Protocols (Revision 4). Reference No.: EMA/540136/2009 [Internet]. 2018 [cited 2020 Feb 26]. Available from: http://www.encepp.eu/standards_and_guidances/checkListProtocols.shtml
56. STROBE Statement. STROBE checklists. 2007 [cited 2020 Feb 26]. Available from https://www.strobe-statement.org/index.php?id=available-checklists
57. Principles relating to processing of personal data. Regulation (EU) 2016/679 of the European Parliament and of the Council of

27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC, OJ L 119, 4.5.2016, Article 5(1)(b).
58. Safeguards and derogations relating to processing for scientific research. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC, OJ L 119, 4.5.2016, Article 89(1).
59. General requirement for consent to general purpose. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC, OJ L 119, 4.5.2016, Recital 33.
60. European Commission Directorate-General for Health and Food Safety. Question and Answers on the interplay between the Clinical Trials Regulation and the General Data Protection Regulation [Internet]. 2019 [cited 2020 Feb 26]. Available from: https://ec.europa.eu/health/sites/health/files/files/documents/qa_clinicaltrials_gdpr_en.pdf
61. Use of de-identified personal data. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC, OJ L 119, 4.5.2016, Recital 26.
62. General requirements for broad consent. Revised Common Rule. 45 Code of Federal Regulation 46.116. [cited 2020 Feb 29]. Available from: https://www.ecfr.gov/cgi-bin/text-idx?SID=1e1c2b5b18e1f4cc624695c38e6057f2&mc=true&node=pt45.1.46&rgn=div5#se45.1.46_1116
63. Office for Human Research Protections. Revised Common Rule Q&As [Internet]. 2018 [cited 2020 Feb 26]. Available from: https://www.hhs.gov/ohrp/education-and-outreach/revised-common-rule/revised-common-rule-q-and-a/index.html
64. Definition of 'covered entities'. HIPAA Privacy Rule. 45 Code of Federal Regulation 160.103. 2019 [cited 2020 Feb 29]. Available from: https://www.ecfr.gov/cgi-bin/text-idx?SID=1e1c2b5b18e1f4cc624695c38e6057f2&mc=true&node=pt45.2.160&rgn=div5#se45.2.160_1103
65. General requirements for authorization. HIPAA Privacy Rule. 45 Code of Federal Regulation 164.508. [cited 2020 Feb 29]. Available from: https://www.ecfr.gov/cgi-bin/text-idx?SID=b35279ee7bdca503d9385c7e6c09b078&mc=true&node=pt45.2.164&rgn=div5#se45.2.164_1508
66. Waiver of authorization. HIPAA Privacy Rule. 45 Code of Federal Regulation 164.512(i)(1)(i). [cited 2020 Feb 29]. Available from: https://www.ecfr.gov/cgi-bin/text-idx?SID=b35279ee7bdca503d9385c7e6c09b078&mc=true&node=pt45.2.164&rgn=div5#se45.2.164_1512
67. Use of de-identified information. Revised Common Rule. 45 Code of Federal Regulation 46.104(d)(4)(ii). [cited 2020 Feb 29]. Available from: https://www.ecfr.gov/cgi-bin/text-idx?SID=1e1c2b5b18e1f4cc624695c38e6057f2&mc=true&node=pt45.1.46&rgn=div5#se45.1.46_1104
68. Use of de-identified information. HIPAA Privacy Rule. 45 Code of Federal Regulation 164.502(d)(2). [cited 2020 Feb 29]. Available from: https://www.ecfr.gov/cgi-bin/text-idx?SID=b35279ee7bdca503d9385c7e6c09b078&mc=true&node=pt45.2.164&rgn=div5#se45.2.164_1502



Author information

Hyunjoo (Mary) Kim is a pharmacist-turned-medical writer who has a great interest in big data. After graduating with an MSc in molecular pharmacology, she worked at Pfizer Korea as an outcomes research and real-world data specialist. In December 2017, she started her career as a medical and regulatory writer at Clinipace Korea.

Zuo Yen Lee graduated with a PhD in biology from ETH Zurich. She found passion in medical writing after a decade of training in academic research and the veterinary diagnostic industry. She is a medical and regulatory writer at Clinipace Taiwan since January 2017. She enjoys every bit of it and is always looking to broaden her horizons.

Chow Lih Yew has been a medical and regulatory writer at Clinipace Singapore since 2018. She enjoys following the latest trends in clinical research and drug development. Before this, she gained over 10 years of experience in academic research in Japan and Switzerland, where she earned her PhD in molecular biology.

Leveraging standardised data in response to the novel coronavirus outbreak

Derk Arts,[1] **Lauren McGill**[2]

[1] Castor EDC, Hoboken, NJ, USA
[2] Freelance Medical Writer, Ontario, Canada

Correspondence to:

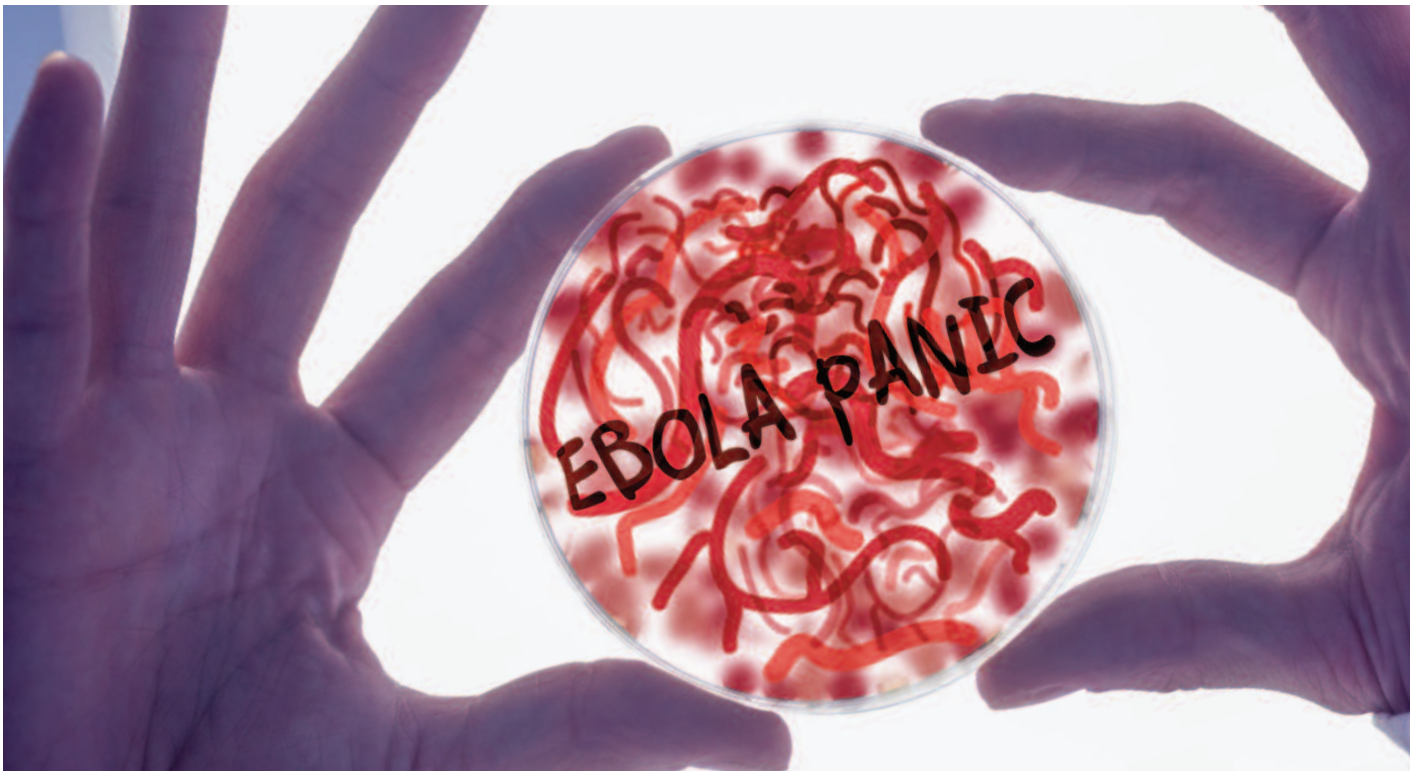
Derk Arts
221 River Street
9th floor
Hoboken, NJ 07030
USA
derk@castoredc.com

Abstract

As we are confronted by a new global health epidemic in the form of COVID-19, the challenges and opportunities of global data sharing come into sharp focus. Due to significant data collection and sharing issues during the 2013–2016 Ebola outbreak, the WHO recently called for improvements before the next public health emergency occurred. While it is too early to quantify the role of standardised data collection and sharing in containing the spread of COVID-19, it is possible to identify some of the data tactics used as part of the medical community's initial response.

Solid data are the best basis for public health action during an unfolding health emergency. Currently, the world is facing just such a crisis. On January 30, 2020, WHO declared the new contagious coronavirus COVID-19 a Public Health Emergency of International Concern. The virus spread rapidly from a single Chinese city through the entire country in just 30 days.[1] Over the following weeks, the world watched as an increasing number of countries reported confirmed cases, triggering government action and worldwide panic. While challenging, this latest global health crisis may prove key in the testing and implementation of new ways to collect, share, and aggregate data.

Castor vs. COVID-19



Lessons from the past

During an outbreak, data are our most valuable assets in the race to effective containment and finding a cure or vaccine. Effective governmental responses are only possible when there is accurate, real-time data available to base decisions on. At the beginning of the COVID-19 outbreak, WHO set clear expectations for better data collection and sharing than during previous outbreaks. In their statement on the second meeting of the "International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV)", WHO declared:

> *As this is a new coronavirus, and it has been previously shown that similar coronaviruses required substantial efforts to enable regular information sharing and research, the global community should continue to demonstrate solidarity and cooperation, in compliance with Article 44 of the IHR (2005), in supporting each other on the identification of the source of this new virus, its full potential for human-to-human transmission, preparedness for potential importation of cases, and research for developing necessary treatment.*[2]

Sadly, the need for "regular information sharing and research" was starkly highlighted during the disastrous handling of data during the 2013–2016 Western African Ebola virus epidemic. In hindsight, it's clear that there were many contributing factors to the difficulties encountered during that health emergency.[3]

Firstly, large pools of existing data from previous Ebola studies, a disease first discovered in 1976, had not been fully disseminated. Research had been conducted, but much of it was never published. When the recent severe outbreak triggered a rush to a cure, incomplete data led researchers down erroneous paths, wasting valuable time. During the multi-year outbreak itself, data were hap-haz-ardly collected and not standardised. There were large communication failures between affected countries and an unwillingness to share information with each other. Finally, data that were actually collected during the outbreak was not always standardised and was often shared inefficiently, with some researchers hesitating to share any data before they were ready for publication in an academic journal.

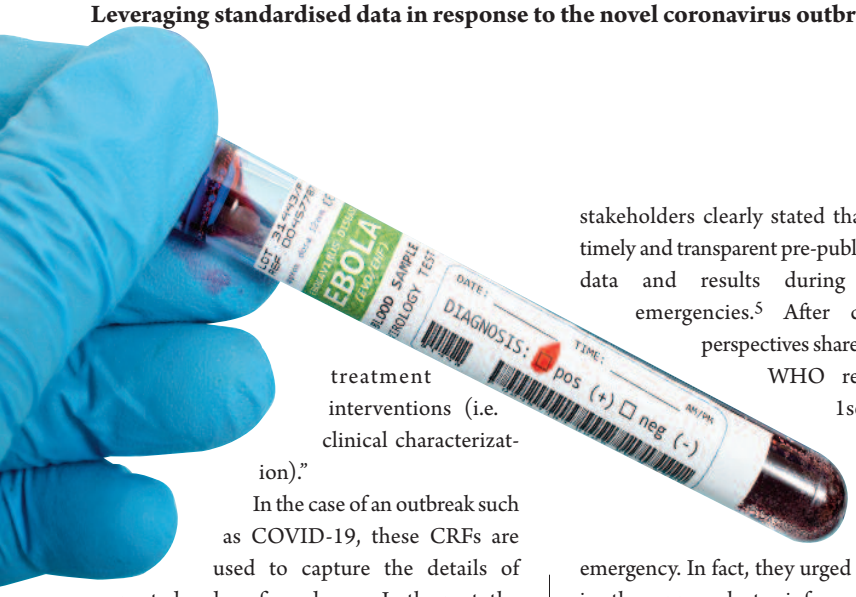
The full impact of the delays in gathering and sharing data during the Ebola outbreak may never be quantifiable. What we do know for certain is that we must rise to the challenge of effective data collection and sharing during current and future outbreaks.

The need for standardised data

In an epidemic, early data are key. It comes with a caveat though: data are only valuable if they are standardised. Amidst any epidemic, the goal is to assemble large amounts of accurate and usable data as quickly as possible. Swaths of data are used to help identify the causative agents; investigate and predict disease spread; define diagnostic criteria; and evaluate treatments and methods to contain further spread. Standardising data means using international recognised terminology for health concepts (e.g., SNOMED or LOINC) to annotate data, or in the very least, capture data in an agreed-upon data model so data from multiple research projects can be pooled and analysed in unison.

The full impact of the delays in gathering and sharing data during the Ebola outbreak may never be quantifiable.

In line with this, as the current COVID-19 outbreak began, WHO quickly provided technical guidance on how to conduct useful early investigations and provided the Global 2019-nCoV Clinical Characterization Case Record Form (CRF).[4] The CRF was "intended to provide member states with a standardized approach to collect clinical data in order to better understand the natural history of disease and describe clinical phenotypes and



treatment interventions (i.e. clinical characterization)."

In the case of an outbreak such as COVID-19, these CRFs are used to capture the details of suspected and confirmed cases. In the past, the reports were paper-based. More recently, medical researchers have used a combination of paper CRFs and electronic CRFs (eCRFs). But with a contagion as fast-moving as COVID-19, speed is key. Therefore, using eCRFs is the best way to hasten the aggregation of clinical data from around the world and accelerate the work of researchers.

In order to contribute to the global response, the medical data capture platform Castor (https://www.castoredc.com/) is providing free access to its platform to support non-profit COVID-19 research projects. As such, it is sharing eCRFs that were built according to the WHO CRF standard. Researchers can start their study or registry in less than an hour, ensuring they capture high-quality data to help drive the global research effort. At the time of publishing, 300 COVID studies haven been created, more than 200 are live and over 5 million data points have been captured on COVID-19 related projects. Further, Castor is supporting the largest global randomised "mega-trial" the find treatments for COVID-19.

The urgency of data sharing

In the post-mortem analysis of the Ebola outbreak response, failure to share relevant data in a timely way has been identified as one of the major hindrances to mounting an effective response. Although the outbreak was eventually contained, lack of data sharing and communication breakdowns delayed acknowledgement of the outbreak's severity and a coordinated response.

Ultimately, the deficiencies of data-sharing mechanisms during the Ebola outbreak became a catalyst for change. In September 2015, WHO held a consultation called "Developing Global Norms for Sharing Data and Results during Public Health Emergencies", where international stakeholders clearly stated that there must be timely and transparent pre-publication sharing of data and results during public health emergencies.[5] After considering the perspectives shared at this meeting, WHO released no-non-1sense recommendations for data sharing during any public health emergency. In fact, they urged "a paradigm shift in the approach to information sharing in emergencies, from one limited by embargoes set for publication timelines, to open sharing using modern fit-for-purpose pre-publication platforms". Of course, such a massive shift requires buy-in from researchers, journals, funders, and others.

WHO went on to prescribe open data sharing as the default response to a public health emergency, declaring that sharing results should be standard practice during a public health emergency. Their recommendations included strong encouragement of sharing epidemiological and population-based data. They warned of the great risks associated in withholding data and results arising from analyses of that data. As we have seen with COVID-19, the risks associated with epidemics are often not shouldered by a single community or nation but rather by the whole planet. The price of data hoarding is simply too high to be allowed any longer.

In line with its own recommendations, WHO immediately began working with its own networks of researchers and other experts to coordinate global work on surveillance, epidemiology, modelling, diagnostics, clinical care, and treatment of COVID-19. It also launched a Global 2019-nCoV Clinical Data Platform.[6] This allows member states to contribute anonymised clinical data, widening the breadth and depth of data collected. Data sharing also went well beyond WHO's own platform. For example, the release of full viral genome sequences through open databases resulted in the development of rapid and reliable diagnostic tests within weeks.[7]

Another issue that was brought forward during the meeting was the unacceptability of non-disclosure of clinical trial data related to research and development in the context of public health emergencies. During a health crisis, decision-makers rely on information disseminated through peer-reviewed journals and accompanying online data sets. Outside of public health emergencies, 12 months is generally considered an acceptable time frame from study completion to public disclosure. However, in an emergency context, WHO recommended that this time frame should be greatly shortened.

Medical journals have responded by taking bold steps to make information available right away. Some of these steps were first tested out during the Zika virus epidemic, when relevant manuscripts were posted online in open collections within 24 hours of submission while undergoing peer review.[8] This trend has continued during the COVID-19 outbreak. In their position statement regarding sharing data during a public health emergency, *The New England Journal of Medicine* states: "Funder signatories will require researchers undertaking work relevant to public health emergencies to establish mechanisms to share quality-assured interim and final data as rapidly and widely as possible, including with public health and research communities and the World Health Organization."[9] Additionally, most major medical journals are providing free access to any and all articles relevant to COVID-19. These changes represent a major shift away from waiting many months, even years, before making highly relevant data accessible to all interested parties.

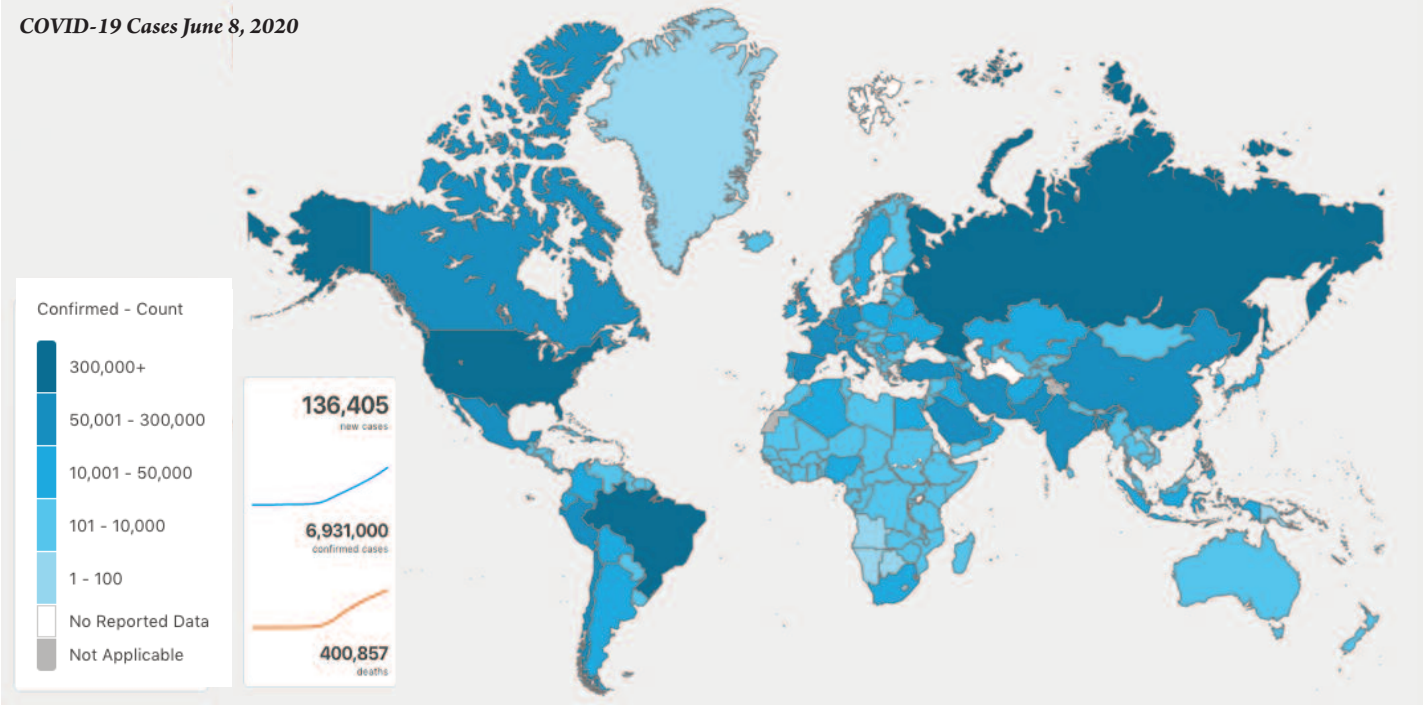
> Epidemics and pandemics spread fast. They do not wait for clinical trials or academic journals to publish results.

Conclusion

Epidemics and pandemics spread fast. They do not wait for clinical trials or academic journals to publish results. In order to contain the current crisis, our scientific communities must leverage the power of data through standardised datasets. With this latest public health emergency, we have an opportunity to get it right. We can accelerate the discovery of cures through cooperation and collaboration. The best way to save lives is to share meaningful data in realtime.

Conflicts of interest

Derk Arts is the CEO of Castor EDC, a health tech company that produces medical data capture software.



References

1. Wu Z, McGoogan JM. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. JAMA. 2020;323(13):1239–42.
2. World Health Organization. Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV). [cited 2020 Mar 2]. Available from: https://www.who.int/news-room/detail/30–01–2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov)
3. Georgetown University Medical Center. Data Sharing during the West Africa Ebola Public Health Emergency: Case Study Report. [cited 2020 Mar 8]. Available from: http://www.glopid-r.org/wp-content/ uploads/2019/07/data-sharing-during-west-africa-ebola-public-health-emergency-case-study-report-georgetown.pdf
4. World Health Organization. Novel coronavirus (COVID-19) – rapid version. [cited 2020 Mar 2]. Available from: https://www.who.int/docs/default-source/coronaviruse/who-ncov-crf.pdf?sfvrsn=84766e69_2
5. World Health Organization. Developing global norms for sharing data and results during public health emergencies [cited 2020 Mar 8]. Available from: https://www.who.int/medicines/ebola-treatment/blueprint_phe_data-share-results/en/
6. World Health Organization. Coronavirus disease (COVID-19) technical guidance: The Unity Studies: Early Investigations Protocols. [cited 2020 Mar 8]. Available from: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/early-investigations
7. Sheridan C. Coronavirus and the race to distribute reliable diagnostics. Nature Biotechnol. 2020;38:382–4.
8. Dye C, Bartolomeos K, Moorthy V, Kieny MP. Data sharing in public health emergencies: a call to researchers. Bull World Health Org. 2106;94:158.
9. New England Journal of Medicine. Zika Virus. [cited 2020 Mar 8]. Available from: https://www.nejm.org/zika-virus

Author information

Derk Arts, PhD, MD, is an epidemiologist and the founder of Castor. As CEO, he's responsible for Castor's growth by staying one step ahead of the field, defining long-term strategy, and building an awesome team.

Lauren McGill is a Canadian freelance medical writer with more than 15 years in various healthcare roles. She believes that storytelling can turn data into compelling narratives that promote health literacy and provider buy-in.

Biostatistics, data management, and medical writing:

A multidisciplinary approach to the development of the CTD integrated summaries

Shiri Diskin,[1] **Tanya du Plessis,**[1]
Ofra Barnett-Griness[2]
Bioforum Group

[1] Bioforum Group, Ness Ziona, Israel, and Bloemfontein, South Africa
[2] Statistical consultant to Bioforum Group

Correspondence to:

Shiri Diskin, PhD
VP, Medical Writing
Bioforum Group
3 Golda Meir St
Ness Ziona, Israel
Shiri.Diskin@Bioforumgroup.com

Abstract

Analyses of integrated databases of efficacy and safety are a Food and Drug Administration (FDA) requirement. They are very useful in evaluating the safety and efficacy data gathered in multiple clinical studies. However, their utility is dependent upon the quality of the studies and the data gathering methods, which affect the quality of the data. It also depends on a scientifically sound strategy for pooling and analysis of the data, and finally, on the adequate reporting of results. Early involvement of professionals from the data management and biostatistics fields can facilitate the development of valuable integrated summary of safety (ISS) and integrated summary of efficacy (ISE) through implementation of study design and data management strategies that are geared toward pooling of data from multiple studies. Medical writers should also join the process early to acquire the knowledge and understanding required for reporting the data in an accurate and meaningful way.

Introduction

A clinical development programme of a pharmaceutical product is designed to collect information that is pertinent to the evaluation of its benefit-to-risk ratio. At later stages of the programme, when preparing for submission, the task at hand is to understand the picture that arises from all available data. In this article, we focus on the rigorous analysis and presentation of integrated clinical safety and efficacy data gathered from prospective, interventional, sponsor-initiated studies from the perspectives of the data management (DM), programming, biostatistics, and medical writing (MW) functions.

Clinical safety and efficacy data from a full programme can be presented in two main ways:

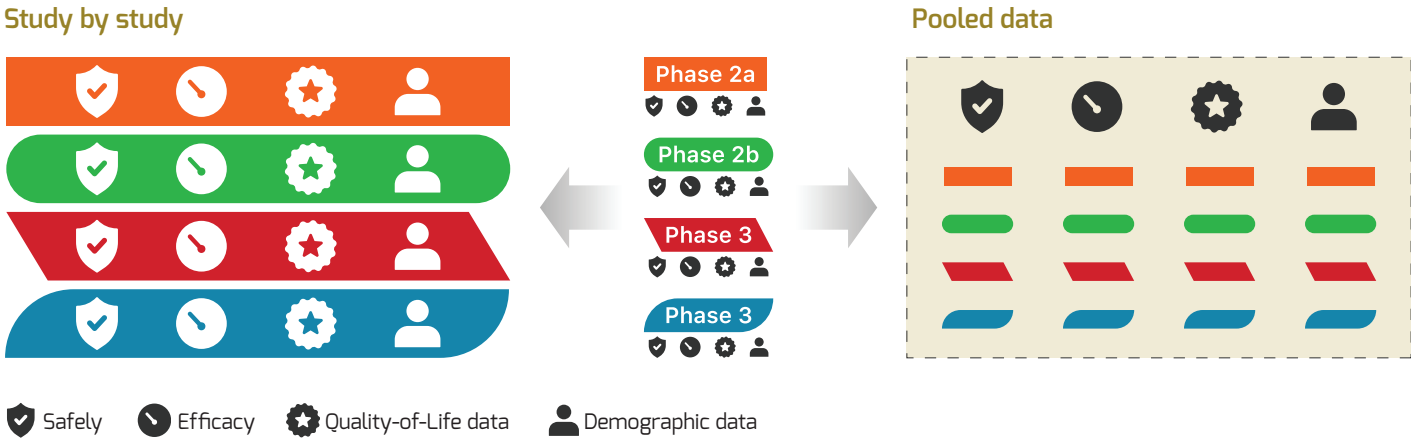


Figure 1. A schematic illustration of data analyses in summary documents
Data can be presented as separate studies side by side, allowing the comparison of specific types of data between studies. Data can also be presented in a pooled manner, referring to the group of studies as a single dataset, providing the benefit of a large sample size.

by study, or by data module. The latter can be done following pooling of data from a few studies into an integrated database that is used as one large study (see Figure 1). This approach can provide valuable tools for understanding the "big picture" as well as addressing specific clinical issues with the product, examples are provided later in this article.

Regulatory requirements

The International Council for Harmonisation (ICH) guideline M4E (R2)[1] *Common technical document for the registration of pharmaceuticals for human use – Efficacy*, refers to three levels of detail of clinical efficacy and safety data presentation (See Figure 2):

- The clinical overview that includes the overviews of efficacy and safety (Modules 2.5.4 and 2.5.5). These are required across ICH countries and are intended as concise and critical analyses of clinical data pertinent to the evaluation of efficacy and safety of the medicinal product in the intended population, focusing on interpretation and discussion.
- The summary of clinical efficacy (SCE) and summary of clinical safety (SCS) (Modules 2.7.3 and 2.7.4). They are required across ICH countries and provide detailed factual summarisation of all data relevant to efficacy and safety in the intended patient population and may include a summary of the results from integrated databases.

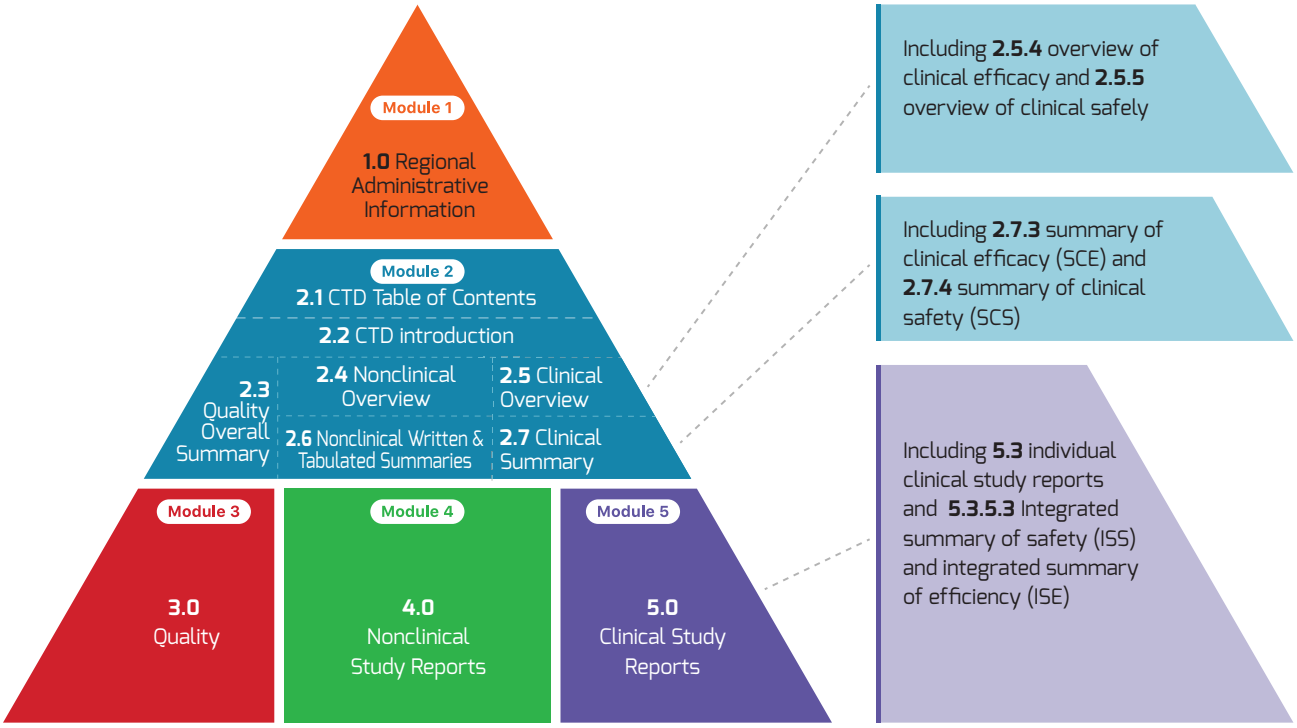


Figure 2. A schematic representation of the hierarchical structure of the Electronic Common Technical Document

- The detailed reports of individual clinical studies written in accordance with the ICH E3 guideline[2] (Module 5.3.5). In addition, FDA requires reports of analyses of data from more than one study, the integrated summary of efficacy (ISE) and the integrated summary of safety (ISS) (Module 5.3.5.3) that provide a detailed description and presentation of the results obtained from integrated safety and efficacy databases.[3]

The data management perspective

There are a few methodologies for the gathering of data in clinical studies. A solid data strategy plan across all studies involved with as much consistency as possible regarding data capture and cleaning allows for a streamlined analysis of data, a reduced need for retrospective reviews and processing.

When initiating a Phase 2a study, many sponsors are uncertain whether their product would be eligible for marketing approval submissions, and at times, are not yet adequately funded. For these reasons, instead of forming a long-term programme-wide data strategy, they opt for "minimum essential" data capture and management plans by using cheap, less reliable, and inconsistent methods for data capture such as paper case report forms or excel sheets.

The FDA requires adherence to specific database design standards for database submissions[4] (such as Clinical Data Interchange Standards Consortium [CDISC] Study Data Tabulation Model [SDTM]). When planning Phase 2b studies, companies are more likely to choose electronic data capture (EDC) systems, however, due to a lack of awareness or lack of resources, FDA-required standards are not always taken into account.

For the purpose of pooling and cross-study analyses, data have to be available in a consistent format (for example, all adverse event [AE] data should be coded in the same version of the MedDRA dictionary[5]). Phase 3 studies are usually designed in collaboration with DM experts with submission in mind. An adequate EDC system with CDISC-compliant data capture is likely to be chosen. Moreover, when a Phase 3 programme includes more than one study, the structure of the studies, the duration of treatment, visit schedules, and data collection of safety and efficacy variables are all planned in a consistent manner allowing for standardisation of data capturing and cleaning. Legacy data should be processed to achieve the same standard. Legacy data that were captured on paper are manually inputted into an EDC system retrospectively with minimal resources for data cleaning and resolving queries. Electronic legacy data that were captured in a format that is not CDISC-compliant must be converted. Taken together, the retrospective processing incur additional costs and time that can be minimised if a data strategy for integration and submission is implemented early on in development.

The biostatistics perspective

It is essential to apply statistical considerations when forming the data integration strategy. The analytical strategy should include the following elements:

- The objectives for integration
- The regulatory guidelines
- Which studies to pool
- The outcomes and time points
- The statistical methods

Regulatory guidelines and integration objectives

The following relevant guidelines can be used in establishing a strategy for the integrated summaries or pooling of data across studies:

- "Summarising the Clinical Database" in the ICH E9[6]
- "Meta-Analyses of Randomised Controlled Clinical Trials to Evaluate the Safety of Human Drugs or Biological Products (2018)", draft FDA guidance[7]
- "Integrated Summary of Effectiveness" (2015), an FDA guidance.[8]

The ISE aims to provide insights beyond those observed in individual clinical trials.[8] Individual studies are presented to demonstrate the claimed effects, and if applicable and appropriate, a statistical combination (pooling) of results may then be considered.

Generally, studies are pooled for two main purposes: to achieve a greater power and increase precision and to assess the drug effect in demographic or clinical subpopulations when there are too few subjects in each individual study to support meaningful conclusions.[8] However, many of the pooled analyses are exploratory in nature and are designed to probe the data for trends across studies, e.g., in disease-specific subgroups.[8]

The draft FDA guidance[7] primarily focuses on meta-analyses with predefined hypotheses designed to confirm a suspected safety risk associated with a drug rather than on exploratory meta-analyses. As such, it is a source for detailed and scientifically rigorous discussions on important considerations when pooling study data for regulatory purposes. It is a valuable resource even when there is no formal predefined safety hypothesis to check.

Which data to pool

When considering which data to pool, the first step is to list all clinical studies with their critical design characteristics, and their respective roles in the development programme.

Importantly, any study included in pooled data will be evaluated and discussed thoroughly as an individual study.

The aim of data integration is to provide a valid description of expected safety and efficacy in the target population, and its usefulness is dependent on individual trials with high-quality data. Thus, the first principle is **Quality over Quantity.**

The application of this principle to efficacy data integration, involves limiting the candidate list to individual studies that are considered "adequate and well-controlled" and serve as the basis for establishing efficacy claims.[6] When applied to Safety integration, findings from a limited set of trials, selected with careful attention to trial and data quality and the intended use of the product, can yield a more informative view of product safety than a broader set of trials that includes trials of poor quality.[7]

The candidate studies are likely non-homogenous, and thus, it should be assessed

A single integrated database can be planned if all candidate studies are deemed similar enough for integration. Otherwise, multiple data pools can be proposed for groups of similar studies. For example, pooling studies using a certain active control or studies with long-term follow-up.

At later stages of the programme, when preparing for submission, the task at hand is to understand the picture that arises from all available data.

whether they can be meaningfully pooled with respect to important elements such as:

- the population, including demographic or clinical factors (such as age range and disease severity);
- the exact indication being evaluated (per study objectives and inclusion/exclusion criteria);
- relevant trial design factors, such as the control group (if any), treatment and follow-up duration, allocation ratio, and collected endpoints.

The degree of variability tolerated with respect to each factor may differ between the integrated datasets for efficacy and safety. For instance, the analysis of AEs may be appropriate in an integrated dataset that includes subjects with different disease severities, but it may be harder to draw efficacy conclusions from such a varied dataset.

A single integrated database can be planned if all candidate studies are deemed similar enough for integration. Otherwise, multiple data pools can be proposed for groups of similar studies. For example, pooling studies using a certain active control or studies with long-term follow-up.

Healthy volunteer studies should not be included in integrated datasets for either efficacy or safety because they assess a population distinctly different from the target patient population and are commonly much shorter. Those studies will be analysed separately and may be pooled as a distinct safety cohort.

Statistical methods can be applied to adjust for important differences between patients and studies that will form the integrated database (see below).

Endpoints and time points

When combining efficacy data, the focus should be on the prespecified primary endpoints (defined for the confirmatory Phase 3 studies). However, when important outcomes are common to all studies (even when the primary endpoints differ), analyses of such outcomes can provide an important assessment of consistency.[8] An example provided in the ISE guidance[8] is a series of studies in which an important variable was assessed at multiple time points, and an analysis of the results obtained at a common time point can be shown, even when the time point for the primary analysis differed among studies.

Unlike efficacy endpoints, safety endpoints are generally standard (usually AEs, laboratory data and vital signs), thus, pooling of all safety outcomes can be expected. Still, the collection timepoints of these measures may not be uniform and studies may vary in treatment duration and follow-up. The use of common time points shared by all the studies and the statistical handling of differences in follow-up duration may facilitate integration of these measures across studies.

The statistical analysis methods

Efficacy analyses are mainly comprised of a pooled treatment effect estimation, by comparing treatment and control groups using appropriate statistical models. A regression model can be applied to the integrated database, the same as would be used for each endpoint in the individual confirmatory study, and further stratified for the study factor. The pooled treatment effects are accompanied by assessment of the homogeneity of the treatment effect across studies, by contrasting the study-specific treatment effects and testing study-by-treatment interaction. A forest plot presenting individual study results and pooled effect is often provided. Thus, the approach taken to derive the pooled effect is to treat the integrated database as a single large trial (see Figure 1), while accounting for study variability (and maintaining randomisation within each study) through stratification.

The statistical analyses for the common safety outcomes are descriptive in nature yet require

special attention. AEs for example, are generally analysed in individual clinical study reports using crude percentages (number of patients with events divided by number of treated patients).

However, naive pooling of the safety database by treating it as one large study and calculating crude percentages for each treatment group, may result in bias when trials employ different randomisation allocations.[7] An illustration of this bias, termed Simpson's paradox is provided in the FDA draft guidance:[7] in the example, the risk for a specific safety event was identical for the treatment and control groups in each of the trials. Thus, analysis of the individual studies would result in the safety event not being a concern for the product. However, the risk was not the same across the trials: in one of the trials, in both the treatment and the control groups, the risk was higher than in others. This study employed a 4:1 allocation ratio (treatment: control), thus simple pooling of trials enriched the treatment group with high-risk patients leading to a biased overall result of increased risk. Statistical solutions for this bias (employing stratification by study and weighting approach) are provided in Chuang-Stein and Beltangady 2011.[9] Contrasting the results from a pooled analysis with that of each specific study can help understand whether a bias has occurred – this substantiates the importance of presenting individual study results as well.

We have assumed throughout that subject-level data is available, as opposed to only trial-level summary measures, for which meta-analytic statistical methods are available.

The medical writers' perspective

MWs are responsible for taking the data and all the background materials and turning them into a coherent narrative that conveys the current knowledge about the efficacy and safety of the product. Therefore, the MW should get involved early in the process of planning submission documents. The MW, whether in-house or outsourced, should be familiar with the clinical studies of the programme and understand their distinct characteristics as well as the shared characteristics that make them eligible for pooling. It is also highly important to understand the evaluation methods used in each study and how each result contributes to the overall claim.

The presentation of both efficacy and safety in the documents should focus on the sought indication in the target patient population, taking care to include the types of information that the regulator is specifically interested in.

Preparation for writing the submission documents should be based on three classes of documents:

1. Sponsor-submitted documents, including Clinical Study Report (CSRs) from earlier studies in the programme and any submission documents from previous programmes of the same product
2. Documented communications with regulatory authorities
3. Templates and guidance provided to sponsors and reviewers [10,11]

Together, these background materials can help the MW understand the company message, how the knowledge about the product has evolved over time and the agreements reached with the different regulators relevant to global submission. The presentation of both efficacy and safety in the documents should focus on the sought indication in the target patient population, taking care to include the types of information that the regulator is specifically interested in.

The documents in Module 2.7 and in 5.3.5.3 (see Figure 2) generally follow the same outline, however, the 2.7 summary documents are limited

in length (up to approximately 200 pages) and cannot delve into the same level of detail as the ones in 5.3.5.3 that can be thousands of pages long. Thus, it may be advisable to start by writing the ISS and ISE, and then summarise the most important information, taking care not to "cherry pick" favourable results. Reduction in volume can be achieved by fewer methodological details about the pooling and integration, as well as ample use of cross-references from the summary documents to corresponding sections of the 5.3.5.3 documents and to source tables and listings.

In terms of project management, the summary documents and the CSRs of the pivotal studies on which the marketing application relies, are likely to be written concurrently. Thus, a submission should be written by a team of MW, that should maintain very frequent communication to ensure that the messages and focus are consistent across documents. The MW team should plan for multiple cycles of cross-review of submission documents. It may be reasonable to assign a leader of efficacy documents (ISE, SCE) and leader of safety documents (SCS, ISS).

Summary

In conclusion, for valid and informative results of drug safety and efficacy, it is recommended to design studies and their data capture and management strategies with integration in mind. It is important to select trials for pooling with careful attention to trial design and data quality, and to combine selected studies using appropriate statistical methods while being careful with naive data pooling. Like the DM and biostatistics professionals, MW should get involved early in the planning for submission, allowing them to get acquainted with the pivotal points of product information and devise a project management plan.

Disclaimers

The opinions and suggestions presented in this article are based on the authors' cumulative experience and do not refer directly to any specific global submission situation.

Conflicts of interest

The authors declare no conflicts of interest.

References

1. M4E(R2) – Common technical document for the registration of pharmaceuticals for human use – Efficacy [Internet]. EMA/CPMP/ICH/2887/1999 Committee for Human Medicinal Products; 2016. Available from: https://www.ema.europa.eu/en/documents/scientific-guideline/ich-m4e-r2-common-technical-document-registration-pharmaceuticals-human-use-efficacy-step- 5_en.pdf
2. Structure and content of clinical study reports, July 1996 [Internet]. CPMP/ICH/137/95 European Medicines Agency; Available from: https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-3-structure-content-clinical-study-reports-step-5_en. pdf
3. Guidance for Industry Integrated Summaries of Effectiveness and Safety: Location Within the Common Technical Document. Food and Drug Administration; 2009.
4. Providing Regulatory Submissions In Electronic Format – Standardized Study Data Guidance for Industry [Internet]. Food and Drug Administration; 2014. Available from: https://www.fda.gov/media/82716/download
5. MedDRA | [Internet]. [cited 2020 Apr 29]. Available from: https://www.meddra.org/
6. ICH E9 statistical principles for clinical trials [Internet]. European Medicines Agency; 1998. Available from: https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-9-statistical-principles-clinical-trials-step-5_en.pdf
7. Meta-Analyses of Randomized Controlled Clinical Trials to Evaluate the Safety of Human Drugs or Biological Products Guidance for Industry – DRAFT GUIDANCE [Internet]. Food and Drug Administration; 2018. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/meta-analyses-randomized-controlled-clinical-trials-evaluate-safety-human-drugs-or-biological
8. Integrated summary of effectiveness-Guidance for industry [Internet]. Food and Drug Administration; 2015. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/integrated-summary-effectiveness
9. Chuang-Stein C, Beltangady M. Reporting cumulative proportion of subjects with an adverse event based on data from multiple studies. Pharm Stat. 2011;10(1):3–7.
10. Day 80 assessment report – Clinical template with guidance rev.02.20 (DOCX/116.3 KB) (updated) [Internet]. Available from: https://www.ema.europa.eu/en/human-regulatory/marketing-authorisation/assessment-templates-guidance#day-80-and-day-120-assessment-report-templates-(containing-guidance)-section
11. Good Review Practice: Clinical Review of Investigational New Drug Applications [Internet]. Food and Drug Administration. Available from: https://www.fda.gov/media/87621/download

Author information

Shiri Diskin, PhD, has been a medical writer since 2009, and is currently working as head of medical writing at Bioforum group, a global data-focused clinical research organisation. Shiri teaches medical writing courses to medical doctors and to biopharma companies.

Tanya du Plessis, M.Med.Sc, CCDM, and PMP, has over 16 years industry experience in data management. Tanya has led many strategies for customised data delivery solutions focusing on high-quality data delivered in a timely manner, while driving numerous innovation teams for optimal data delivery.

Ofra Barnett-Griness, PhD, has been a biostatistician for more than 20 years, working with pharmaceutical industry and academia in the design and analysis of randomised clinical trials and real-world data studies. Ofra is a statistical consultant to Bioforum group.

The new value of clinical data in Europe

Jasminka Roth
The Tao of Excellence, Schaffhausen, Switzerland

Correspondence to:
Jasminka Roth
The Tao of Excellence
Oberberg 2
8222 Beringen, Switzerland
+41 / 76 585 52 00
jasminka.roth@taoexcellence.ch

Abstract
The EU's new Medical Devices Regulation and In-vitro Diagnostic Device Regulation has integrated and reinforced the regulatory requirements for pre- and post-market clinical trials and positioned them as vital to ensuring the safety and performance of a medical device on the market. These regulatory changes combined with greater access to clinical data and the technology that is now available on the market, enable manufacturers to analyse and even use clinical data in new ways. Smart algorithms can now evaluate with speed and ease a product's clinical data to benchmark it in the context of the global market for the very first time. This meta-analysis, driven by artificial intelligence, will in turn redefine the way manufacturers approach the design of clinical trials to make them more effective, efficient, reduce waste, improve product performance and safety, discover new markets, innovate devices and secure a faster pathway to funding and reimbursement.

High clinical data requirements
New regulation, technological advances, and artificial intelligence are redefining healthcare systems and medical industries around the world, but the medical devices industry is undergoing one of its most radical changes to date. Whilst producing and collecting clinical data have become more high-profile internationally, the most significant increase in regulatory requirements for clinical data has been prescribed by the European Commission's new Medical Devices Regulation (EU) 2017/745 (MDR). The effective date of the regulation has been postponed by 1 year to May 26, 2021, because of the global impact of the COVID-19 pandemic.

The MDR requires manufacturers to supply clinical data for devices both pre- and post-market as evidence of the product's performance and safety. The standards for clinical data are also higher and will apply to more medical devices than ever before in an effort to harmonise the quality of both data and devices throughout the EU. This new level of demand for clinical data presents a significant challenge for manufacturers as clinical trials are one of the most expensive and time-consuming aspects of launching a product on the market.

Small and medium-sized enterprises (SMEs), which represent 95% of medical device manufacturers in the EU,[1] are likely to be hit hardest by this new financial burden as they tend to have more limited resources, a higher dependency on a single market, and are usually more reliant on the success of a single product. As such, SMEs represent both the largest majority of manufacturers in the EU and the demographic that will need clinical evaluations to do far more than just be a vehicle to meet regulatory requirements.

The high cost and waste involved in clinical data
Generating clinical data incurs a wide variety of high and inescapable costs. Specialist knowledge, supplies, and facilities, as well as the length of time it takes to recruit for and complete a study, all contribute to the high cost of clinical trials. In 2010, a Stanford study analysing clinical data from the FDA estimated that the average cost of bringing a 510(k) product from concept to market was $31 million, but more than 77% percent of the cost (approximately $24 million) was consumed by regulatory and FDA-related activities. Similarly, the cost of pre-market approval averaged at $94 million, 80% of which was spent on the regulatory stages required to bring a medical device to market launch.[2]

SMEs represent both the largest majority of manufacturers in the EU and the demographic that will need clinical evaluations to do far more than just be a vehicle to meet regulatory requirements.

In addition to the often-soaring costs of clinical trials, investigations into clinical research practices over the years have consistently revealed a high volume of waste, with some experts finding as much as 85% of medical research is wasted globally and avoidably.[3] "This waste arises from the multiplicative effects at different stages of research: over 50% of research is not published; over 50% has avoidable design flaws; and over 50% is unusable or incompletely reported, or both."[4] Compounding this waste are issues in methodology, design bias, the misuse of statistical methods, poor reproducibility, and insufficiently rigorous studies, which in the end all amount to wasted research, resources, and time. And whilst real-world data is used for purposes such as marketing and sales we could regard it as a waste

that it is not utilised for regulatory purposes, or only very limitedly so.

Disjointed regulatory procedures waste research, resources, and time

Traditionally, pre- and post-market have been different bodies, working in separate timeframes, concentrating on different aspects of the market. It is this division between their regulatory requirements that drives the disjointed perceptions and processes that waste resources, opportunities, and time.

Pre-market, randomised clinical trials (RCTs) tend to be designed with just the regulatory requirements for a product submission in mind. Questions that relate to the clinical perspective, such as the post-market clinical follow-up (PMCF), the health tech assessment (HTA), and the quality-adjusted life year equation – all of which ultimately identify what products will receive funding and reimbursement from the healthcare system, are therefore often overlooked as they are not required for product certification. Instead, these questions are often deferred, increasing the likelihood that more data will be required later on, meaning more work and cost for the manufacturer and a slower path to funding and reimbursement.

EU MDR streamlines regulatory practices to optimise clinical data

The EU's MDR now regulates clinical investigations, integrating pre- and post-market regulatory requirements and raising requirements and standards. The MDR, for instance, requires manufacturers to support their product claims pre and post-market via post-market surveillance including the PMCF. It therefore makes sense for clinical studies to collaborate with market access studies – and vice versa – to better predict a product's economic viability as well as its safety and performance on the market.

The MDR reflects the trajectory of development in the wider world. An increasing emphasis and greater reliance on data and technology. Investors, health authorities, health professionals, and patients are demanding more conclusive clinical data, market transparency, and ultimately better and more agile health systems. This has been thrown into far sharper relief by the current COVID-19 crisis, highlighting how important data accuracy, transparency, and technology are to generating insights that enable us to implement actions with confidence, speed, and agility.

> How can the financial burden of clinical data and new technologies deliver more than the regulatory requirements for performance and safety?

Manufacturers therefore need to ask themselves how they can optimise the return on such heavy investments. How can the financial burden of clinical data and new technologies deliver more than the regulatory requirements for performance and safety?

The answer actually lies in the connection between the two. It is through clinical data and

advances in technology that manufacturers have the opportunity maximise the potential of clinical investigations and their evaluations.

Greater access and tech will redefine the use of clinical data

Globally, governments, policy, and regulation are increasing access to clinical data to improve market transparency, safety, and the economics involved in running healthcare systems. Primary registries, such as the FDA's ClinicalTrials.gov (which hosts over 300,000 clinical trials) and the European database on medical devices (EUDAMED) coming up with the MDR, are databases that grant manufacturers, healthcare professionals, patients, and the public levels of access to clinical data. These directives, combined with the real-world data collected by private registries, supply manufacturers with large datasets they can use to draw insights for market analysis and their business.

Technology, machine-learning, and smart algorithms are also revolutionising the way we use clinical data, shifting it from being the preserve of a highly professional sector to more public and commercial use.

For the first time in fact, the medtech industry is experiencing an excess of data generated by both traditional RCTs and new tech solutions such as health apps, wearables, sensors, and software. Produced through partnerships with tech companies, such as Zimmer Biomet and Apple,[5] and Novartis and Microsoft,[6] this new tech increasingly empowers patients to contribute clinical data independently and report the effects of a device or treatment in near-real-time.

Technology, machine-learning, and smart algorithms are also revolutionising the way we use clinical data, shifting it from being the preserve of a highly professional sector to more public and commercial use. Deep 6 AI, for example, uses artificial intelligence to locate appropriate candidates for clinical trials in "minutes instead of months"[7] by using data analysis drawn from ICD-10 codes, unstructured clinical data, pathology reports and operating notes, thereby lowering the cost of clinical investigations.[8] This initiative attracted $17 million in funding in 2019.[7]

In addition to this, COVID-19 has brought to the fore how important and valuable clinical data and health apps are to healthcare. Launching an effective contact tracing app, for instance, like the application program interface (API) solutions that Apple and Google are collaborating on,[9] is being considered as key to halting the spread of the virus and supporting a transition to the new normal after a lockdown. The desire for clinical information is also not restricted to professionals in the sector. The public want access too, as demonstrated by the publicly available COVID-NMA website, which includes a data visualisation overview of the COVID-19 clinical trials being carried out around the world to "monitor in real-time any new evidence that becomes available" and "identify gaps and deficiencies of existing evidence early enough and with an aim to help prioritizing and optimizing future research".[10]

Driving the evolution of meta-analysis

However, whilst registries and databases make clinical data more publicly available, the data itself is not necessarily actionable. Industry expertise is needed to be able to read the information recorded in a clinical trial through meta-analysis. Technical, time-consuming, and performed manually, this process can be cumbersome and doesn't usually allow for lateral insights beyond the regulatory requirements of a clinical evaluation. Clinical data are also recorded in different ways depending on the individual study format and database. So, whilst technology, machine-learning, and smart algorithms can enable us to draw insights from large datasets at speed, the data structure and quality itself still needs to be improved. Initiatives like the EUDAMED are helping to structure this data better and enable its use even further. This is the beginning of a significant shift in how we use clinical data and approach regulatory requirements, and a marked evolution in meta-analysis.

A year ago, we set out to create a tech collaboration to develop an algorithm that would provide manufacturers with better insights into their own clinical data. A smart algorithm is able to recognise, match, and analyse the different variety of data across clinical trials, databases, and languages at a speed that far outpaces traditional methods of analysis. Identifying gaps and deficiencies in a product's clinical data for regulatory requirements or otherwise benchmarking clinical data against existing RCTs.

Network meta-analysis powered by artificial intelligence can evaluate a manufacturer's clinical data to see whether their study is likely to meet the expectations of health authorities. It can

demonstrate how to improve a clinical trial or product for a more competitive edge, illuminate a pathway to funding and reimbursement, highlight a new market, or even identify a gap in the market for a potential innovation. The speed and accuracy of AI-driven meta-analysis also enables manufacturers to quickly and easily identify where to allocate their resources and improve the efficiency and efficacy of clinical trials through comparative analytics, reducing wasteful research and potentially speeding up the time it takes to certify a device.

Again, COVID-19 is starkly highlighting the vital importance of such algorithms if we are to draw real insight from the wealth of data being produced worldwide. Clinical data production is increasing rapidly. Our own algorithm has recorded a 500% increase in the number of RCTs between early March and late April, totalling 443 clinical trials across 57 countries with almost 300,000 patients enrolled. It is these vast datasets combined with the algorithms that can collect, read and analyse the information that are producing the insights we now rely on to shut down entire countries, as well as to reopen them.

Optimise the value of clinical data

Through AI meta-analysis, manufacturers now have the opportunity to question their regulatory processes and ask how they might benefit their business beyond the clinical evaluation. One of the most obvious points of which is to reassess how clinical trials could be designed to include additional endpoints that can be used downstream. For example, putting questions like the HTA or PMCF alongside pre-market RCT questions can produce data that helps to focus or influence post-market research or marketing initiatives. This frees up resources and potentially smooths and speeds up the pathway to funding and reimbursement. It also provides data that can be used to benchmark a product in the market that can help it to improve or become more competitive.

Integrating the design of pre- and post-market clinical trials will not only help manufacturers meet the EU's regulatory requirements but also optimise the resources that they are being required to invest. New technology and AI-driven meta-analysis allows manufacturers to assess clinical data with far greater speed and precision that can help them to both complete their clinical evaluation and transcend the traditional boundaries between pre- and post-market. It can benchmark products in the market, evaluate clinical trial design for regulatory purposes and reveal gaps and opportunities for innovation. AI-driven meta-analysis can discover new markets, new products, improve devices, and define a funding and reimbursement strategy. It can open our imagination and clarify our vision, enabling us to see how we can shape a more sustainable, inclusive, and healthier future.

AI-driven meta-analysis can discover new markets, new products, improve devices, and define a funding and reimbursement strategy.

Conflicts of interest

The author declares no conflicts of interest.

References

1. The European Medical Technology Industry – in figures 2019. MedTech Europe, 2019. p. 20.
2. Makower J. Meer A, Denend L. FDA Impact on U.S. Medical Technology Innovation: A Survey of Over 200 Medical Technology Companies. PwC, 2010. p. 7.
3. Chalmers I. Glasziou P. Avoidable waste in the production and reporting of research evidence. Lancet. 2009;374:86-9.
4. Glasziou P. Chalmers I. Research waste is still a scandal—an essay by Paul Glasziou and Iain Chalmers. BMJ. 2018;363:k4645.
5. Farr C. Apple hopes the Apple Watch can help patients recover faster from knee and hip replacements. CNBC. 2018 Oct 15.
6. Novartis. Novartis and Microsoft announce collaboration to transform medicine with artificial intelligence. 2019 [cited 2019 Oct 1]. Available from: https://www.novartis.com/news/media-releases/novartis-and-microsoft-announce-collaboration-transform-medicine-artificial-intelligence.
7. D'Onfro J. Startup Deep 6 Lands $17 Million To Use AI To Help Find Patients For Clinical Trials. Forbes. 2019 Nov 25.
8. Deep 6 AI. How it works. Available from: https://deep6.ai/how-it-works/
9. Apple Newsroom. Apple and Google partner on COVID-19 contact tracing technology. 2020 [cited 2020 Apr 10]. Available from: https://www.apple.com/newsroom/2020/04/apple-and-google-partner-on-covid-19-contact-tracing-technology/
10. covid-nma.com. Available from: https://covid-nma.com

Author information

Jasminka Roth is a longstanding quality and regulatory expert in the healthcare, pharmaceutical, and medtech industries. Jasminka founded the company The Tao of Excellence in 2015, and the algorithm "Dragonfish" in 2019. She has a MSc and BSc in biochemistry and biotechnology from ETH Zürich and worked at the University of Sydney and University Hospital of Bern, Inselspital.

Unlocking the potential of patient data through responsible sharing – has anyone seen my keys?

Patrick Cullinan[1], **Liz Roberts**[2]

[1] bluebird bio, Inc, Cambridge, MA, USA
[2] UCB Biosciences, Inc, Raleigh, NC, USA

Correspondence to:

Patrick Cullinan
bluebird bio
60 Binney Street,
Cambridge, MA 02142 USA
pcullinan@bluebirdbio.com

Abstract

The sharing of individual participant-level clinical data is now an almost routine extension of the clinical study life-cycle, and increasingly a vital element of leveraging real-word data. Responsible clinical study data sharing of appropriately consented and de-identified participant-level data and associated clinical documents is an expectation of key research funders, journal editors, pharmaceutical trade associations, regulators, ethics committees, and government entities sponsoring research. Furthermore, patients increasingly support expanded data sharing to help spur innovation and maximise the utilisation of data gathered during clinical studies. Finally, rapidly and appropriately leveraging real-world data to support and validate clinical research data and to facilitate responses to emerging public health emergencies lends greater importance and urgency to finding better ways to unlock and share health data. This article provides an overview of the current state of participant-level data sharing in clinical research and a discussion of the opportunities that exist to better navigate barriers to access whilst respecting the data privacy rights of study participants. This article describes our collective journey through the data sharing ecosystem, looking to further unlock the value of study participant data to drive new discoveries.

Background

Data obtained through clinical research are fundamental to advancing the field of medicine and to improving the health and well-being of patients. The data underlying such research have historically remained securely in the custody of the data generator (in most cases a pharmaceutical or academic study sponsor) and access was highly limited. Recently there have been increasing calls from patient groups, advocacy organisations, journal editors, pharmaceutical and biotechnology trade associations, regulators, and others in the scientific community for the responsible sharing of study patient-level data-sets and/or study documentation, to provide greater transparency and propel research innovation through secondary data reuse. Additionally there have been calls by many, including the World Health Organization (WHO), to accelerate and extend these data sharing paradigms to speed up the collection and dissemination of data during public health emergencies, a need exemplified during the recent Ebola outbreaks in West Africa and the COVID-19 pandemic.[1] It is envisaged that by harnessing the statistical power of large data-sets the broader scientific community can embrace the "big data" revolution, including utilising machine learning and artificial intelligence to spur new frontiers in data analytics and data collaboration.[2] Recently, all data generators have been further incentivised to share data by journal requirements to make their data sharing plans public as a pre-requisite commitment for publications (such as those aligned to the International Committee of Medical Journal Editors position and associated with PLoS One journals).[3,4] In addition, in 2016 the FAIR (Findable, Accessible, Interoperable, Reusable) Guiding Principles for Scientific Data Management and Stewardship[5] were published to provide guidelines to data generators, to improve the findability, accessibility, interoperability, and reuse of data which has further helped to propel data sharing, especially from the academic perspective. As such, all clinical research data generators have experienced increasing requirements and calls to establish mechanisms to responsibly share the clinical study data they produce.

This article provides an overview of patient-level data sharing with a focus on clinical trial data sharing, a discussion of key limiters and enablers that require greater attention to optimally unlock the value of patient-level data, and aspects of data sharing that have yet to be fully harnessed to further drive new discoveries.

Data sharing process

Patient-level data sharing refers to the process whereby data providers accept requests from academic and non-academic researchers for access to data and supporting documentation from formal clinical trials. Figure 1 provides an overview of the common steps of the data sharing request process that generally occur, although the process may vary based on the data provider or access platform.

Discoverability

Before a requester can make a data sharing request they need to know what data are available and understand key data characteristics. As such, there is a need for data providers to make public their data sharing policies and to communicate the availability of studies.

With this in mind, the trade association principles for responsible clinical study data sharing require sponsors to make public their data sharing policies and provide a mechanism to receive Research Proposals (RPs). In addition, ClinicalTrials.gov recently implemented an IDP (individual patient data) sharing plan section of the register that must be completed by each sponsor as part of study registration to improve the discoverability of data. As such, most pharmaceutical and biotechnology sponsors have made their data sharing policies and processes public, and increasingly data generators of all types are making their data sharing plans public at the time of study registration and publication. In most cases, pharmaceutical data providers have joined consortia or created stand-alone portals that specifically list studies for sharing to aid discoverability. Together, these measures have

dramatically increased the discoverability of clinical studies available for data sharing RPs and are helping to spur a new era of transparency and data-driven innovation.

Applying for access

Data access requires submission of a robust RP by a requester (usually a researcher on behalf of a broader research team that includes a statistician or suitably qualified data-analytics professional [e.g., health economist]). The RP requests specific studies and notes other data that the researchers will seek to aggregate or otherwise include in their analysis. In many cases, RPs also include data management and statistical analysis plans that outline precisely how they will manage and use the requested data, as well as detailed publication plans and conflicts of interest statements.

Review, approval, contracting, and access

In most cases, RPs are initially reviewed by the data provider for completeness and alignment to an organisation's data sharing policies, the study informed consent, and other legal bases for sharing (e.g., consistent with the European Union [EU] General Data Protection Regulation [GDPR]).[6] Requests are subsequently reviewed by an independent review panel (IRP) to assess the scientific merit and other aspects of the request (e.g., conflicts of interest and researcher qualifications). The manner in which IRPs are utilised (e.g., as the primary review panel or as an appeal panel for sponsor-rejected requests), their role in review, and the degree of independence varies. Upon RP approval researchers and/or their institution must sign a data sharing agreement (DSA) specifying the data access conditions and licences that are being granted. These agreements include a commitment to protect the privacy of study participants and the confidentiality of data provider information, and detail other obligations and rights associated with data access.

Once the DSA is executed, access to anonymised data and de-identified documents are provided, in most cases, via a secure cloud-based research environment. Data protections seek to minimise the risk of participant/patient re-identification and release of company confidential information. The protected data are generally provided for a defined period of time (usually 1 to 2 years), although extensions are possible.

Data sharing landscape

The rapid evolution of data sharing has resulted in the development of a complex and often non-interoperable landscape of data sharing platforms and research environments. This overview provides a high-level landscape summary of the types of data sharing systems and a summary of key platforms that have developed, although it is not intended to represent an exhaustive list. Rather, it is intended to provide a sense of the types of platforms that have developed. For specific information, policies, and processes relating to a specific sharing mechanism, the reader should refer to the applicable data sharing portals or provider websites.

Pharmaceutical study sponsor data sharing

Although ad hoc and fit-for-purpose data sharing has been occurring for some time in the pharmaceutical industry, large-scale and coordinated data sharing implementation in a broader sense gathered momentum in response to the establishment of trade association principles for responsible data sharing in 2014.[7] Early adopter companies developed mechanisms to share data through the establishment of portals to accept requests, IRPs to adjudicate access, and secure data sharing research environments. These early efforts to share data led to the creation of ClinicalStudyDataRequest.com (CSDR) and The YODA Project (Yale Open Data Access) portals.[8,9] Other pharmaceutical entities have created similar partnerships to facilitate data sharing, for example, Duke Clinical Research Institute has partnered with pharmaceutical sponsors and others to facilitate access via Duke's Supporting Open Access for Researchers (SOAR) platform.[10] While there are distinguishing differences between these portals (for example, some aspects of SOAR emphasise curation and data harmonisation), they essentially follow the

same major steps outlined in Figure 1 (study listing, proposal review, IRP approval, DSA, research conduct). One limitation of these platforms is that there has been little interoperability of the research environments established for these entities, although some recent efforts have been made to permit researchers to request data across these and other systems.

While these consortia/academic-supported data sharing platforms have made rapid progress, the majority of pharmaceutical and biotech sponsors share data via stand-alone portals (ranging in complexity from proposal submission gateways to more simple online forms or email request systems) and utilising various company-specific IRPs and data access approaches.

Another pharmaceutical sponsor data sharing arena is related to pre-competitive data sharing intended to permit collaboration amongst sponsors to spur more efficient and effective clinical development. Examples of such pre-competitive data sharing include the IQ (Innovation and Quality) consortium of pharmaceutical and biotechnology companies who share (largely) technically-focused pre-clinical and early clinical data to identify new science, technology, and regulatory engagement pathways.[11] Another is the DataCelerate platform developed to support sharing amongst TransCelerate and BioCelerate member companies.[12]

While the extent of growth in the area of data access has been rapid and impressive, the proliferation of platforms has resulted in data access mechanisms that have limited interoperability and are inefficient and difficult to navigate from a researcher perspective. Efforts are underway to create greater opportunity for cross-platform access and improve the efficiency of the process overall.

Non-profit data sharing portals

While pharmaceutical sponsors have been expanding efforts to directly share data with each other and with independent researchers, non-profit entities have been entering the data sharing space, both in their capacity as funders and through the creation of data sharing infrastructure to further facilitate access, lower the threshold for entry, and imparting further independence to the process. An important non-profit active in this space is the Wellcome Trust, which has been a leader in developing and enforcing data sharing requirements for its funded research and has also served to support the development of both CSDR and Vivli (see below), primarily by supporting the management of IRPs on these platforms.[13]

Direct efforts by non-profit organisations (often associated with specific funders or patient organisations) to support data sharing are largely

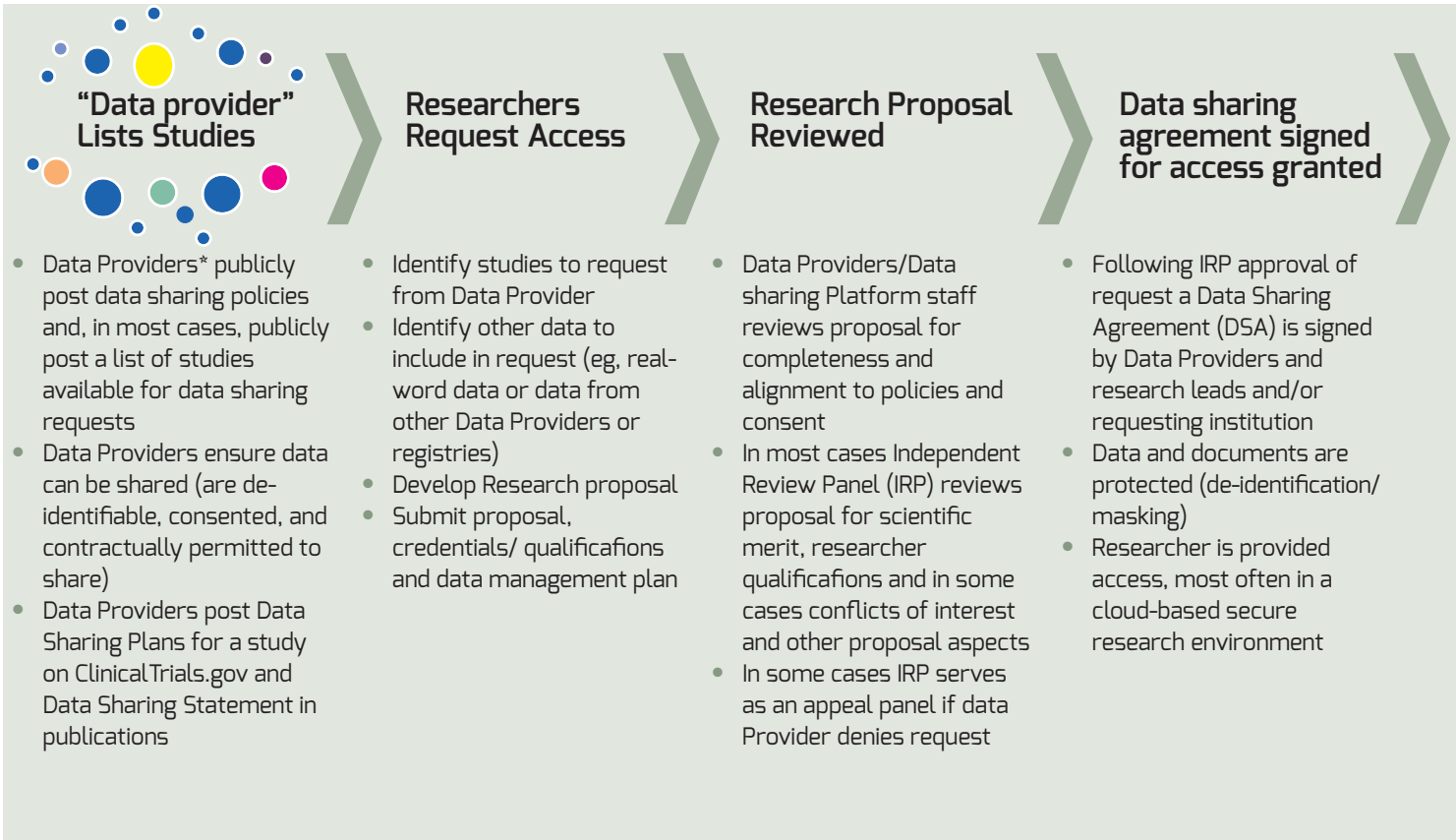


Figure 1. Common data sharing request steps

Not all data providers or data sharing mechanisms require all steps or in some cases may include extra steps (such as an appeal process for denied access requests).

*Data Provider is a broad term intended to encompass all data generators including both academic and pharmaceutical/biotech sponsors and other data generators (e.g., non-profit research entities). It also includes data requested from data stewards who manage acquired data that they did not generate (such as through an acquisition, merger, or via a data donation).

disease- or therapeutic-area specific. An important early example of such sharing is the oncology-specific Project Data Sphere, which was created by *CEO Roundtable on Cancer's Life Sciences Consortium* and which initially focused on the sharing of downloadable comparator-arm data from oncology studies, although this platform also has the capacity to host data in a secure environment or more recently via Vivli (see below).[14]

Another important non-profit data sharing entity is Vivli, which is a data sharing platform that seeks to serve as a neutral broker between data providers, data accessors, and the wider data sharing community (that includes pharmaceutical, academic, non-profit, and public-private data).[15] Vivli has developed a global data sharing and analytics platform that seeks to span all disease areas, facilitate interoperable data sharing across a range of data providers, and intends to expand to create disease-specific communities and add capability to support data protection processing (de-identification) to lower the barrier to entry for smaller entities and academic data providers.

Public-private partnerships

The efficiency and impact of data sharing can be maximised when focused curation and harmonisation of data-sets occurs. As such, data sharing is increasingly an important part of large public-private partnerships (PPPs), which are large initiatives coordinated by governmental (public) and industry/academic (private) entities and created through shared public and private funds. These partnerships generally focus on specific public health priorities and, in doing so, can bring substantial resources and organisational infrastructure to accelerate innovation through enhanced data sharing and other methods.

In Europe, the Innovative Medicines Initiative (IMI) is a partnership between the EU and the European Federation of Pharmaceutical Industries and Associations to address a range of important healthcare research topics. More and more of IMI's projects include data sharing/aggregation aspects focused on enabling access and innovative new research relating to specific diseases.[16] In the US, similar PPPs are funded by the Foundation for the National Institutes of Health with the support of the Pharmaceutical Research and Manufacturers of America and are called Accelerating Medicines Partnerships (AMPs).[17] AMPs seek to accelerate research in a range of disease areas and to support biomarker development. Another PPP model in the US is coordinated by the Critical Path Institute which is a non-profit PPP involving the FDA that aims create new data, measurement, and methods standards through the aggregation of data to spur new innovation in a pre-competitive consortium model via their platforms.[18]

Other data sharing

Patient-level clinical study data sharing is increasingly supplemented by data from other sources. Data from patient registries and patient data aggregation projects (e.g., the UK biobank and the US National Institute of Health "All of Us" campaign), as well as real-world data from electronic health records and other sources such as wearable devices, will increasingly be leveraged by researchers to supplement or compare against clinical study data.[19, 20] Importantly, while the ability to leverage such sources may improve the power of data sharing, such data have limitations in terms of its quality, uniformity, and standardisation. In addition, combining such data sources with clinical study data may represent an increased risk to participant re-identification as the possibility of linking and identifying patients across larger and more diverse data-sets increases.

> By harnessing the statistical power of large data-sets the broader scientific community can embrace the "big data" revolution, including utilising machine learning and artificial intelligence to spur new frontiers in data analytics and data collaboration.

Data yet to be fully unlocked

While there have been substantial advances in the field of data sharing, as outlined below, there are several types of data that are not optimally being shared.

Rapid data sharing during public health emergencies

For the most part, the data sharing mechanisms outlined previously take a deliberate approach to data sharing that seek to responsibly account for the privacy and consent of participants and protect the intellectual rights of researchers and sponsors to protect confidential information and data rights. Such "responsible data sharing" therefore takes a methodical approach that can be potentially time consuming and it may not be possible to share certain data due to confidentiality, privacy, and other legal limitations. Such sharing is not suitable for rapid data dissemination as is needed during a public health emergency. Insufficient timely access to reliable data severely hampers epidemiological tracking to the spread of disease and efforts to coordinate control and implement treatment responses and research. Our collective deficiencies in rapidly collecting and sharing basic scientific data (such as viral gene sequences), real-world data (such as sharing infection rates, patient symptoms, disease

Research conducted and project closed

- Researcher conducts project per proposal analysis plan
- Researcher request extension if needed or amends project (with appropriate approvals)
- Researcher notifies Data Providers/IRP of any new safety signal, intellectual property or publications
- Access to the project is closed per DSA
- Manuscript submitted for publication

trajectory, epidemiological data, and outcomes to treatment protocols), and early clinical trial data in a manner that is responsible, timely, accurate, and interoperable, to appropriately inform public health policy was identified during the recent Ebola outbreaks in West Africa in 2014 to 2016, the subsequent emergence of Zika infections, and more recently the COVID-19 pandemic.[21,22] Unfortunately with every outbreak there is an urgent need to establish or re-establish ad hoc mechanisms for expedited data collection, sharing, and publication rather than implementing the utilisation of established standards and systems. An example of an effort to provide a mechanism for such sharing is the recent implementation by the WHO of a "COVID-19 Open" data sharing and reporting protocol, which seeks to provide a mechanism for rapid online publishing of COVID-19 research papers – similar to approaches implemented following the emergence of Zika infections.[1,21,22] These rapid peer-reviewed publications are among the many efforts to share results that the WHO and others have repeatedly attempted to implement during health emergencies, yet clearly there is a need to proactively establish infrastructure and data standards for the collection and sharing of data during health emergencies that can overcome geopolitical, language, and other barriers and support informed scientific research and health-policy decisions through more timely sharing of standardised data.

To this end, at the time of writing, certain pharmaceutical companies, not-for-profit organisations, academia, and health authorities have united across various platforms to explore new ways to collaborate and responsibly share data more promptly. How this sustainably changes the paradigm of data sharing post-pandemic is yet to be seen.

Rare-disease data

Sharing of rare-disease data represents an innovation opportunity if challenges relating to patient privacy can be overcome. Aggregation of rare-disease patient-level data can help overcome the paucity of patients participating in research, for example, by providing historical control data. However, most data providers have policies that exclude sharing any data where the risk of re-identification of patients would be elevated, and as such, do not share data from studies in diseases considered rare. To overcome this issue, the broader scientific community is working to develop advanced data anonymisation and sharing technology (possibly utilising encryption, synthetic data modelled on the actual data, or employing distributed analytic techniques that bring the analytics to the data [rather than sharing the data itself]) and enhanced patient consents that more clearly consent patients by outlining the risks of re-identification and potentially offering patients the option to opt out of sharing, and alternative legal bases for sharing and managing rare-disease patient data.

Genomic data and biospecimen sharing

Genomic data by their very nature are unique to a given individual and so represent immense potential that is limited by privacy concerns. In addition, these data are very sensitive, and their misuse could have implications beyond an individual (e.g., having implications for family members and a potential generational impact). As such, efforts to broadly share genomic data and to tie the sharing of such data to clinical study data or other data sources have been limited, although early examples have emerged (e.g., the Psychiatric Genomics Consortium) and mechanisms are being developed to better enable such sharing.[23] Similarly, biospecimen/sample sharing represents another underutilised data resource that has been limited by concerns related to consent, import/export and privacy regulations, re-identification risk, logistical matters, and lack of clarity related to "ownership" and data-steward responsibility for newly-generated data from the sample. Enabling better utilisation of genomic and biospecimen data in a responsible manner with adequately informed and consented patients and leveraging new technologies to protect patient privacy will be important to unlocking the huge potential of these data sources.

Keys to drive enhanced data sharing

While there has been substantial recent progress towards enhanced FAIR access to participant-level data, there remain substantial barriers that continue to limit access and hamper the efficiency of the ecosystem and medical communication professionals can play an important role in unlocking the data.

One important way in which medical communication professionals can enable data sharing is to consider this topic in the development of informed consents and protocols. Clearly discussing data sharing in these documents can facilitate later data sharing. Informing patients of the sponsor's data sharing plans, possible use of such data, and residual privacy risks associated with sharing of de-identified data can substantially enable future data sharing. Such consent and protocol language can help clarify the legal basis of sharing as it relates to evolving privacy legislation and streamline ethics committee approval.

Medical communicators also play an important role in improving discoverability through including appropriate and clear data sharing plans on clinical trial registers and publications. Furthermore, medical writers can support subsequent data sharing processes by employing lean-writing approaches that minimise the need to redact while still producing documents of high clinical utility. Indeed, making data and documents easier to protect can enhance utility to the research community and can make the sharing process more efficient.

Other important enablers of data sharing efficiency include broadening the use of common data standards and more prospectively releasing information about the structure and contents (data dictionary/metadata) of "to be shared" data. Posting such information along with the listed title and basic metadata would allow researchers to more effectively plan and understand "what they are getting", thus enabling more efficient and successful data sharing.

Conclusions

Data sharing has made immense progress in the past five years yet more can be done to unlock its true potential, especially considering emerging disease challenges that will require data driven solutions. Medical writers are well positioned to be a key contributor to facilitate progress in this space. Making data more discoverable, improving protocol and patient consent language relating to data sharing and the associated residual risks, ensuring clear description of the legal basis for sharing, and improving the timeliness, efficiency, and utility of shared data and documents through lean authoring and writing with privacy protection in mind, can substantially unlock and enable enhanced data sharing.

Disclaimers

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

Conflicts of interest

The authors are employed by bluebird bio (and a bluebird stock owner) and UCB Biosciences, respectively, and either in the context of their current or past employers have supported or participated in some of the data sharing platforms discussed in this paper, particularly ClinicalStudyDataRequest.com and Vivli.org.

References

1. Modjarrad K, Moorthy VS, Millett P, Gsell PS, Roth C, Kieny MP. Developing global norms for sharing data and results during public health emergencies. PLoS Med. 2016 Jan;13(1):e1001935.
2. Institute of Medicine. Sharing clinical trial data: Maximizing benefits, minimizing risk. Washington DC: The National Academies Press; 2015.
3. Taichman DB, Backus J, Baethge C, et al. A disclosure form for work submitted to medical journals: a proposal from the International Committee of Medical Journal Editors. Lancet. 2020 Jan;S0140-6736(20)30187-2.
4. Bloom T. PLOS' New Data Policy: Part Two. In: EveryONE: PLOS ONE Community Blog [internet]. 2014 [cited 2020 Mar 06]. Available from: https://blogs.plos.org/everyone/2014/03/08/plos-new-data-policy-public-access-data/
5. Wilkinson M, Dumontier M, Aalbersberg I, et al. The FAIR guiding principles for scientific data management and stewardship. Sci Data. 2016:160018. doi: 10.1038/sdata.2016.18.
6. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation [GDPR]).
7. European Federation of Pharmaceutical Industries and Associations, Pharmaceutical Research and Manufacturers of America. Principles for Responsible Clinical Trial Data Sharing: Our Commitment to Patients and Researchers. 18 Jul 2013 [cited 2020 Mar 06]. Available from: https://efpia.eu/media/25189/principles-for-responsible-clinical-trial-data sharing.pdf
8. Clinical Study Data Request Consortium data sharing platform [internet]. [cited 2020 Mar 06]. Available from: https://www.clinicalstudydatarequest.com/
9. Yale Open data Access data sharing platform [internet]. [cited 2020 Mar 06]. Available from: https://yoda.yale.edu/
10. Duke Supporting Open Access for Researchers (SOAR) data sharing platform [internet]. [cited 2020 Mar 06]. Available from: https://dcri.org/our-work/analytics-and-data-science/data sharing/
11. International Consortium for Innovation and Quality in Pharmaceutical Development (IQ) data sharing platform [internet]. [cited 2020 Mar 06]. Available from: https://iqconsortium.org/about/
12. DataCelerate data sharing platform [internet]. [cited 2020 Mar 06]. Available from: https://transceleratebiopharmainc.com/datacelerate/
13. Wellcome Trust Data, software and materials management and sharing policy [internet]. [cited 2020 Mar 06]. https://wellcome.ac.uk/grant-funding/guidance/data-software-materials-management-and-sharing-policy
14. Project Data Sphere data sharing platform [internet]. [cited 2020 Mar 06]. Available from: https://projectdatasphere.org/projectdatasphere/html/home
15. Vivli data sharing platform [internet]. [cited 2020 Mar 06]. Available from: https://vivli.org/
16. the Innovative Medicines Initiative Projects and Results [internet]. [cited 2020 Mar 06]. Available from https://www.imi.europa.eu/projects-results
17. Foundation for the National Institute of Health Accelerating Medicines Partnerships [internet]. [cited 2020 Mar 06]. Available from: https://fnih.org/what-we-do
18. Critical Path Institute data access portals [internet]. [cited 2020 Mar 06]. Available from: https://c-path.org/programs/
19. UK Biobank data sharing platform [internet]. [cited 2020 Mar 06]. Available from: https://www.ukbiobank.ac.uk/
20. All of US data access platform [internet]. [cited 2020 Mar 06]. Available from: https://www.researchallofus.org/workbench/
21. Dye C, Bartolomeos K, Moorthy V, Kieny MP. Data sharing in public health emergencies: a call to researchers. Bull World Health Organ. 2016 Mar;94(3):158.
22. Moorthy V, Henao Restrepo AM, Preziosi M-P, Swaminathan S. Data sharing for novel coronavirus (COVID-19). Bull World Health Organ. 2020 Mar;98(3):150.
23. Sullivan PF, Agrawal A, Bulik CM, et al. Psychiatric genomics: An update and an agenda. Am J Psychiatry. 2018;175(1):15–27.

Author information

Patrick Cullinan, PhD, is Senior Director of Medical Writing for bluebird bio, a gene therapy company in Cambridge Massachusetts. In addition to medical writing, Patrick has broad experience leading clinical trial transparency and data sharing teams, and was a former member of ClinicalStudyData Request.com and Vivlii.org steering committees and has supported data sharing planning and initiatives for a range of non-profit and public-private partnerships organisations.

Liz Roberts, MSc., is an experienced Senior Director with almost 30 years' experience within pharmaceutical R&D. Transitioning from a career in biostatistics and with a demonstrated history in the field of clinical data transparency, disclosure, and patient-level data sharing, she provides external thought-leadership to increase the value these activities can bring to people living with severe diseases. Particular areas of interest include value-based transparency, the legal framework supporting data sharing activities, and the co-creation of forward-looking, meaningful policies, and deliverables.

COVID-19 contagion, information, and misinformation

Melvin Sanicas
Takeda Pharmaceuticals
Zurich, Switzerland

Correspondence to:
Melvin Sanicas
Takeda Pharmaceuticals International AG
Thurgauerstrasse 130
8152 Glattpark-Opfikon
Zurich, Switzerland
melvin.sanicas@takeda.com

Abstract
The set of reactions being observed in the current coronavirus outbreak is similar to that in other epidemics: the Severe Acute Respiratory Syndrome, the swine flu H1N1 pandemic, and the Middle East Respiratory Syndrome. Indeed, progress has been made, such as increase in the speed of viral genome sequencing and vaccine development. However, the spread of misinformation also proceeds faster than at any point in history. The world should learn its lessons from this experience and explore appropriate alternatives.

We have seen this happen before – with SARS (Severe Acute Respiratory Syndrome) in 2002, with H1N1 swine flu pandemic in 2009, with MERS (Middle East Respiratory Syndrome) in 2012. A novel pathogen appears, the worst-case scenario is assumed, and some people focus on an existential life-or-death scenario allowing fear and panic to win over logic and reason. Not unexpected. It is not different now with the novel coronavirus, SARS-CoV, and the disease it causes, Coronavirus Disease-2019 (COVID-19).

General facts on SARS-CoV
Scientists understand from the genetic sequencing of the novel virus that it came from animals, as did the other problematic coronaviruses, SARS and MERS. As there was a cluster of early cases linked to a seafood market in Wuhan, which sold and slaughtered live wild animals for food, most people believe the likely source to be wild animals. The similarity of this virus to SARS suggests the source animal to be a bat, but there would have been an intermediary animal carrying the virus prior to infecting humans, which could have been civet cats, bamboo rats, or pangolins, according to a preprint (manuscript that has not undergone peer review) posted by Xiao et al.[1]

After more than 1,400,000 cases of COVID-19 worldwide,[2] what are still unknown about the virus outnumber the known. For example, it is known that the modes of transmission are via respiratory droplets from coughing or sneezing infected persons and contact with contaminated surfaces. There have been reports of airborne transmission in China; however, the World Health Organization (WHO) disagrees that it is a major driver of transmission.[3]

Some infected individuals may not develop any symptoms at all but the symptoms of the majority of the cases closely resemble those of the common cold[3] (dry cough, fatigue, and fever). Some may present with body malaise, sore throat, or nasal congestion. One out of six infected individuals develop difficulty in breathing and in rare instances, the infection may lead to mortality. Data thus far suggest that COVID-19 case fatality risk is around 1% (the figure is higher in Wuhan but lower in Singapore, South Korea, and Italy); this puts it somewhere between the 1957 influenza pandemic (0.6%) and the 1918 influenza pandemic (2%).[4] Current epidemiological results show that elderly persons and those with pre-existing medical conditions are at the highest risk of developing the severe form of the disease.[4]

Among what are not known include details on how this coronavirus causes the disease, interacts with proteins, or responds to seasonal changes and the damages it leaves a patient surviving the disease.[5] Coronaviruses, in general, baffle scientists.

However, given the current growth rate of cases, affected geographies, and high percentage of people who survive rather than die from the disease,[2] there is reason for optimism that the outbreak will taper off and not worsen.

Reaction to outbreak
Unfortunately, this novel coronavirus has spawned an "infodemic" – as coined by the WHO for the collection of theories of conspiracies, unsubstantiated claims, and phony cures surrounding COVID-19.[6]

Social media and some irresponsible health care websites have taken advantage of the predictable pattern of public anxiety. Misinformation (see e.g., Agence Press-France[7]) has become more than a distraction hindering an effective public health response. As they spread faster than the SARS-CoV itself[8] they exacerbate the outbreak by promoting "cures" or prevention methods for coronavirus, which are ineffective, non-evidence based, and may likely be dangerous. In Iran, at least 300 have died from methanol poisoning after an article saying methanol was a cure for COVID-19 became viral.[9]

The public health community and the world of medical science should effectively communicate facts and address feelings as a key part of communication and preparedness, rather than

dismiss them. For example, the message of the Singaporean Head of State did not only include giving concrete facts and actions to be taken, but also recognising people's fear and making people in control of their situation.[10] In fact the Singaporean government has won praise for its response to the outbreak and shown the world that good communication is an essential ingredient in pandemic response and preparedness. With over 1,400 confirmed cases as of April 7, 2020, Singapore has one of the lowest death rates (6 fatalities) in Southeast Asia.[2]

A novel pathogen appears, the worst-case scenario is assumed, and some people focus on an existential life-or-death scenario allowing fear and panic to win over logic and reason.
Not unexpected.

Positive developments surrounding pandemic

COVID-19 has strengthened collaboration among scientists all over the world and triggered a faster working pace among them. Vaccine development is a long and complex process, often lasting 10 to 15 years: from identification of the genome sequence (at least one year) until several phases of clinical trials. With a sense of urgency, scientists were able to sequence the genome of the virus in a matter of weeks.[11]

Moderna, a biotech company based in Cambridge, Massachusetts, has already created an mRNA vaccine against SARS-CoV-2 that encodes a version of the viral spike protein.[12] Apart from Moderna, the Coalition for Epidemic Preparedness Innovations has also been funding development of an mRNA vaccine from CureVac, in Tubingen, Germany, a DNA vaccine from Inovio Pharmaceuticals[13] in Plymouth Meeting, Pennsylvania, and a protein vaccine from a research group at the University of Queensland in Australia.[14] At the end of February 2020, nearly 300 papers had already appeared on preprint repositories (bioRxiv and medRxiv) compared with 261 published in journals. Many scientists now welcome the new way of communicating study results due to the outbreak.[15] However, we should be cautious on how we interpret preprints as these publications have not endured scientific scrutiny from independent experts (peer review).

Measures to be undertaken at individual level

While waiting for antivirals and vaccines to be approved by regulatory authorities and be available widely, one can do the following:

1. Regular washing of hands. Doing so dramatically reduces chances of transmitting or contracting both respiratory and

Table 1. Recommended reliable sources of information on COVID-19

Organisation	Website
World Health Organization	https://www.who.int/emergencies/diseases/novel-coronavirus-2019
US Centers for Disease Control and Prevention	https://www.cdc.gov/coronavirus/2019-ncov/index.html
European Centre for Disease Prevention and Control	https://www.ecdc.europa.eu/en/novel-coronavirus-china
European Commission Coronavirus Response Team	https://ec.europa.eu/info/live-work-travel-eu/health/coronavirus-response_en
European Medicines Agency	https://www.ema.europa.eu/en/human-regulatory/overview/public-health-threats/coronavirus-disease-covid-19
The New England Journal of Medicine	https://www.nejm.org/coronavirus
Johns Hopkins CSSE COVID19 tracker	https://systems.jhu.edu/research/public-health/ncov/

diarrhoea-causing illnesses.[16,17]

2. Getting an influenza vaccination. Although it will not prevent infection from the novel coronavirus, it will lessen the number of people showing symptoms similar to COVID-19 and thus reduce the amount of people competing for hospital care.
3. Avoiding hoarding of masks. Indeed, masks offer some protection (by providing physical barrier). However, public health experts agree that the likelihood of being infected is higher through using the hands in touching things and then touching the face afterwards than not wearing masks.[18] Washing hands is still the most important prevention method. In addition, hoarding masks will diminish supply for medical professionals who actually need them while they treat the sick.
4. Getting facts right. In the digital era, everyone obtains information via the internet. However, care should be taken as dubious messages circulate on *WhatsApp* and other forms of social media. One should rather listen only to the advice of public health bodies, such as the WHO, the US Centers for Disease Control, the European Centre for Disease Prevention and Control, and local health authorities (Table 1).

Conclusion

The COVID-19 outbreak is a complex interplay among the contagion, information, and misinformation. The long-term challenge is how to improve our ability to respond to outbreaks. This is the very same challenge faced during the outbreaks of SARS, H1N1 swine flu pandemic, MERS, and other pandemics even decades before. Some things have already changed such as the advancement in science and technology including speed of information dissemination. However, we need to proactively address misinformation related to an outbreak as it might affect people's behaviour and put them in danger.

Acknowledgements

The author would like to thank Merlin Sanicas and Melissa Sanicas for assistance.

Disclaimers

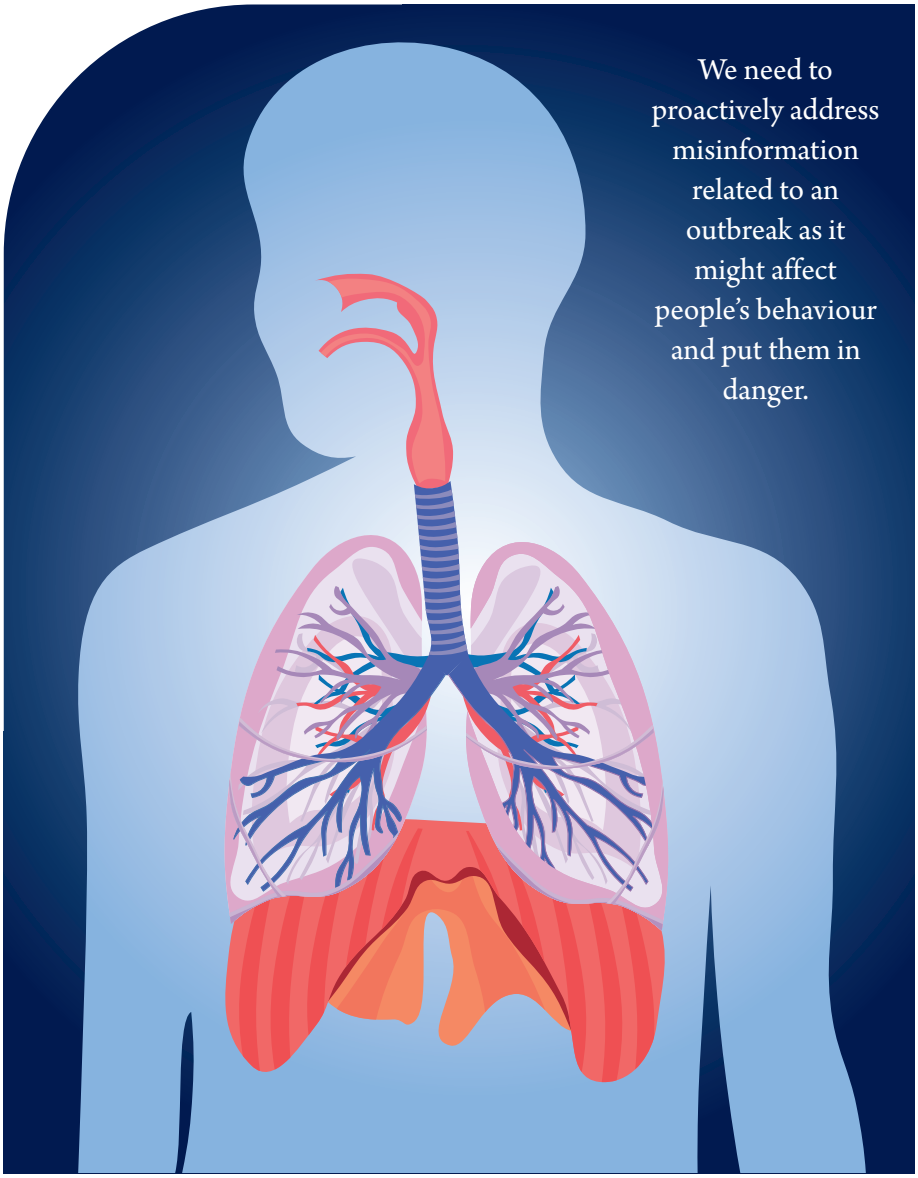
The opinions expressed in this article are the author's own and not necessarily shared by his employer or EMWA.

Conflicts of interest

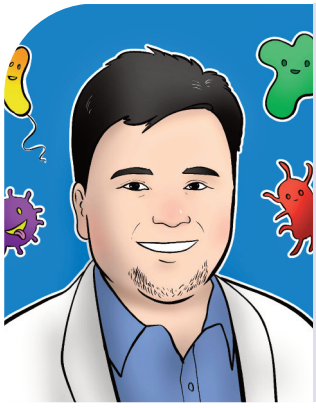
The author is employed by Takeda Pharmaceuticals International AG, a company developing vaccines that tackles problems in public health including dengue, norovirus, Zika, and chikungunya.

References

1. Xiao K, Zhai J, Feng Y, et al. Isolation and characterization of 2019-nCoV-like coronavirus from Malayan pangolins. Preprint. Posted online February 20, 2020. bioRxv 2020.02.17.951335. doi: https://doi.org/10.1101/2020.02.17.951335
2. Worldmeter. COVID-19 Coronavirus Pandemic. 2020 [cited 2020 April 8]. Available from: https://www.worldometers.info/coronavirus/
3. World Health Organization. Q&A on coronaviruses (COVID-19). 2020. [cited 2020 Feb 24]. Available from: https://www.who.int/news-room/q-a-detail/q-a-coronaviruses
4. The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team. The epidemiological characteristics of an outbreak of 2019 novel coronavirus disease (COVID-19) – China, 2020. China CDC Weekly. 2020;2:1–10.
5. Neumann B. What the coronavirus does to your body that makes it so deadly. The Conversation. 2020 Apr 02; [cited 2020 Apr 10]. Available from: https://theconversation.com/what-the-coronavirus-does-to-your-body-that-makes-it-so-deadly-133856
6. Naughton J. Fake news about Covid-19 can be as dangerous as the virus. The Guardian. 2020 Mar 14. [cited 15 March 2020]. Available from: https://www.theguardian.com/commentisfree/2020/mar/14/fake-news-about-covid-19-can-be-as-dangerous-as-the-virus
7. Agence France-Presse. Coronavirus: Russia pushing fake news about US using outbreak to 'wage economic war' on China, officials say. South China Morning Post. 2020 Feb 23. [cited 24 Feb 2020]. Available from: https://www.scmp.com/news/world/russia-central-asia/article/3051939/coronavirus-russia-pushing-fake-news-about-us-using
8. Vosoughi S, Roy D, Aral S. The spread of true and false news online. Science. 2018;359:1146–1151.
9. Roger C. Social media misinformation that led to 300 deaths in Iran claimed that drinking methanol was a cure for COVID-19. Tech Times. 2020 Mar 29. [cited 2020



Mar 30]. Available from: https://www.techtimes.com/articles/248428/20200329/social-media-misinformation-that-led-to-300-deaths-in-iran-claimed-drinking-methanol-was-a-cure-for-covid-19.htm
10. Beaubien J. Singapore wins praise for its COVID-19 strategy: The U.S. does not. NPR. 2020 Mar 12. [cited 15 Mar 2020] Available from: https://www.npr.org/sections/goatsandsoda/2020/03/12/814522489/ singapore-wins-praise-for-its-covid-19-strategy-the-u-s-does-not
11. Yasinski E. Scientists scrutinize new coronavirus genome for answers. The Scientist. 2020 Jan 23. 2020. [cited 24 Feb 2020] Available from: https://www.the-scientist.com/news-opinion/scientist-scrutinize-new-coronavirus-genome-for-answers-67006
12. Keener AB. Just the messenger. Nature Med. 2018;24:1297–1300.
13. Cohen J. With record-setting speed, vaccine makers take their first shots at the new coronavirus. Science. 2020 Mar 31. [Cited 8 Apr 2020]. Available from : https://www.sciencemag.org/news/2020/03/record-setting-speed-vaccine-makers-take-their-first-shots-new-coronavirus
14. Australian Trade and Investment Commission. COVID-19 vaccine project at University of Queensland gets a A$17 million boost. 2020 Mar 25. [cited 30 Mar 2020]. Available from: https://www.austrade.gov.au/international/invest/investor-updates/2020/covid-19-vaccine-project-at-university-of-queensland-gets-a-a-17-million-boost
15. Kupferschmidt K. A completely new culture of doing research: Coronavirus outbreak changes how scientists communicate. Science. 2020 Feb 26. [cited 2020 Mar 30]. Available from: https://www.sciencemag.org/news/2020/02/completely-new-culture-doing-research-coronavirus-outbreak-changes-how-scientists
16. Feltman R. Want to stay healthy? Learn to wash your hands the right way. Popular Science. 2020 Feb 18. [cited 2020 Mar 30]. Available from: https://www.popsci.com/story/health/how-to-wash-your-hands-avoid-illness/
17. Centers for Disease Control and Prevention. Show me the science – why wash your hands? 2020 Sept 17. [cited 2020 Mar 30]. Available from: https://www.cdc.gov/handwashing/why-handwashing.html
18. Khalil S. Face masks 'increase risk' of getting virus. The Chronicle. 2020 Mar 2. [cited 2020 Apr 10]. Available from: https://www.thechronicle.com.au/news/face-masks-increase-risk-of-getting-virus/3957948/



Author information

Melvin Sanicas, physician and scientist, is the Medical Director Clinical Excellence at Takeda Vaccines Business Unit. He is a TED Educator, a contributor for the World Economic Forum, a global assessor for the Royal Society for Tropical Medicine and Hygiene, and a fellow of the Royal Society of Arts.

About this article

I first met Brussels-based lawyer, An Vijverman, at the European Centre for Clinical Research Training Data Transparency Conference in February 2019. The conference brought together the clinical trials industry – sponsors and contract research organisations – regulators, watchdogs, and academics. There were sessions on regulatory aspects; how the data transparency rules are lived; proactive approaches to data transparency; anonymisation; and the future of data transparency. My colleague Tracy Farrow and I spoke about smart-authoring clinical study reports (CSRs) to reduce required effort with data sharing, using CORE Reference in the "proactive approaches" session. An spoke with authority grounded in her legal expertise, adding a new dimension to the "regulatory aspects" session. It was enlightening for me as a medical writer charged with writing CSRs that share data responsibly to hear about the legal aspects to enable me to apply the regulations as intended. An spoke about compliance with the EU Clinical Trials Regulation (CTR), Medical Devices Regulation, and General Data Protection Regulation (GDPR) and how these regulations all impact data transparency.

Hoping to take advantage of her legal brain, I asked An if informed consent forms (ICFs) need to reflect that the data gathered in clinical trials are disclosed through the publication of clinical documents such as CSRs in the EU, and if so how do ICF authors need to adapt standard ICF texts. An was able to explain this in non-legal jargon, that I found easy to navigate, and she kindly agreed to write this up and share it with readers of *Medical Writing* in the following article.

Sam Hamilton

Processing health-related data for scientific research: Is consent an appropriate legitimate ground?

An Vijverman
Lawyer-partner Dewallens & partners, Belgium (www.dewallens-partners.be).

Correspondence to:
An Vijverman
an.vijverman@dewallens-partners.be

Is informed consent needed to process clinical trial data?

Since May 25, 2018, the GDPR[1] has come into application. This means – or should mean – a harmonisation of the rules on data processing throughout Europe. However, the European Member States continue to interpret certain aspects of data processing differently, such as the interpretation of legitimate grounds needed for processing health-related data for scientific research. Not all Member States are aligned as to whether the informed consent of the participant is required to process health-related personal data for scientific research purposes.

What does GDPR say?

Article 6 GDPR lays down the possible legitimate grounds for processing personal data in general. Article 9.1 GDPR further prohibits the processing of health-related data, except if one of the conditions laid down in Article 9.2 GDPR is fulfilled. When processing health-related personal data, the Controllers should ground their processing on one of the legitimate grounds laid down in Article 6 GDPR, as well as on one of the legitimate grounds laid down in Article 9.2 GDPR.

GDPR allows for processing of health-related data without informed consent being given

One of the legitimate grounds laid down in Article 6 GDPR and one of the conditions of Article 9.2 GDPR concerns the (explicit) consent of the data subject (Articles 6.1.a and 9.2.a GDPR). However, Articles 6.1 and 9.2 GDPR also contain other possible legitimate grounds that could be used for processing health-related personal data for scientific research purposes.

So, according to Article 6 GDPR, one could alternatively also justify the processing of personal data for scientific research purposes as the processing that would be necessary for the purposes of the legitimate interests pursued by the Controller (Article 6.1.f GDPR) or because the processing is necessary for the performance of a task carried out in the public interest (Article 6.1.e GDPR).

Furthermore, according to Article 9.2.j GDPR, one could then alternatively justify the processing of health-related personal data for scientific research purposes as the processing that would be necessary for scientific research purposes, provided Article 89 (1) GDPR is also respected and provided the processing is proportionate to the aim pursued; respects the essence of the right to data protection; and provides for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject (Article 9.2.j GDPR). This means that if the processing of health-related data is necessary for scientific research purposes and if all other conditions laid down in Article 9.2.j GDPR are complied with, the consent of the data subject (Article 9.2.a GDPR) will not be required.

Article 89 (1) GDPR further contains the obligation to minimise data when performing scientific research. This means that if scientific research can be performed based on data processing that does not permit or no longer permits the identification of data subjects, one should do so. In other words, data should as far as possible be pseudonymised or – if possible – anonymised when being used for scientific research.

Hence according to the GDPR, the processing

of health-related data for scientific research is possible, without disposing of the informed consent of the data subject, provided the data are as far as possible pseudonymised or anonymised and provided the principle of proportionality and the right to data protection and the fundamental rights and interests of the data subjects are complied with.

Informed consent is not even recommended

I would even go a step further and advise not to ask for informed consent from the data subject to cover the processing of his or her personal data for scientific research purposes, based on another legitimate ground: a consent is only valid if it has been freely given (Article 7 GDPR) and consent should not be regarded as freely given if the data subject does not have a genuine or free choice or is unable to refuse or withdraw consent without detriment (Consideration 42 GDPR). Also, in order to ensure that consent is freely given, consent should not provide a valid legal ground for processing personal data in a specific case where there is a clear imbalance between the data subject and the Controller, in particular where the Controller is a public authority and it is therefore unlikely that consent was freely given in all the circumstances of that specific situation (Consideration 43 GDPR). In the context of scientific research, this means that if a patient decides to participate in scientific research, he or she cannot really freely consent or not with the processing of his or her personal data for that scientific research. Indeed, participating in scientific research *ipso facto* also implies the processing of the data subject's health-related data for the purpose of that scientific research. One may therefore conclude that the consent the patient would give for the processing of his or her health-related data for the purpose of scientific research cannot be given freely (and would thus by definition be invalid). The patient moreover has a subordinate relationship towards the investigator and/or research institution which implies an imbalance of power also seriously complicating free consent.

What does the European Data Protection Board say?

This reasoning has been confirmed by the European Data Protection Board in Opinion 3/2019[2] (Art. 70.1.b).

Legal opinion and advice to the authors of ICFs

I therefore conclude that it is not advisable to use consent as the legitimate ground for processing health-related personal data as this entails the risk that the freely given character of the consent is subject to discussion afterwards, in which case the data cannot (any longer) be processed legitimately. It is better to use instead the legitimate grounds of the necessity for the legitimate interests of the Data Controller/the necessity for performing a task carried out in the public interest (Article 6.1.f or e GDPR) and the necessity for scientific research purposes (Article 9.2.h GDPR).

To avoid any misunderstanding, I confirm that this reasoning is limited to the processing of personal data as part of scientific research but does not in any way influence the rules on performing the scientific research itself. Indeed, the rules on performing scientific research itself – requiring in most cases the informed consent of the participant as to his or her participation in the trial – continue to apply. It essentially concerns the rules laid down in the European Directive 2001/20//EC of the European Parliament and of the Council of April 4, 2001, on the approximation of the laws, regulations and administrative provisions of the Member States relating to the implementation of good clinical practice in the conduct of clinical trials on medicinal products for human use and in European Regulation (EU) 536/2014 of the European Parliament and of the Council of April 16, 2014, on clinical trials on medicinal products for human use, and repealing Directive 2001/20/EC – the latter is expected to come into application during 2020.

References

1. Regulation (EU) 2016/679 of the European Parliament and of the Council of April 27, 2016, on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).
2. Opinion 3/2019 of the European Data Protection Board of 23 January 2019 concerning the Questions and Answers on the interplay between the Clinical Trials Regulation (CTR) and the General Data Protection Regulation (GDPR).

An interview with
Cathal Gallagher
of EMA's Technical Anonymisation Group

About this article

In our daily work in a company or freelance setting, we interact with other relevant functions, typically biostatistical, medical, programming, and data management colleagues, so that we can deliver well-rounded deliverables that take account of multiple perspectives. If we take a higher-level strategic view for the developing regulatory public disclosure (RPD) arena, then we should be talking to professionals outside of our usual networks to ensure we understand and take account of broader perspectives. I engaged with Cathal Gallagher of EMA's Technical Anonymisation Group and asked him the kinds of questions that you as medical writers might, and hope to have contextualised his RPD perspectives to our work in writing clinical documents fit for public disclosure.

Sam Hamilton

Sam Hamilton: I am delighted to be talking to Cathal Gallagher, a member of the EMA's Technical Anonymisation Group (TAG).[1] EMA's TAG is an expert group in data anonymisation, and they aim to help further develop best practices for the anonymisation of clinical reports, in the context of the EMA's policy on the publication of clinical data (Policy 0070).[2] Cathal, you have kindly agreed to tell us more about the TAG and what the work of the TAG means in the context of the work that regulatory medical writers do. Let's start by understanding a little about your career and how you ended up on EMA's TAG.

Cathal: I used to be an IT and Maths high school teacher. In late 2011, I decided that I needed a change and made an effort to retrain myself in SAS programming. SAS software is used to create the tables, figures, and listings (TFLs) that medical writers typically use as source data in their clinical study reports (CSRs). With my newly found skills, I was hired by a small company doing SDTM[3] and ADaM[4] programming. These two Clinical Data Interchange Standards Consortium (CDISC) standards[5] are used to develop clinical trial datasets and are the required standards for both FDA's and Japan's Pharmaceutical and Medical Devices Agency (PMDA) dataset submissions. This small company was bought by d-wise, who were building a piece of software to anonymise clinical trial datasets. I got involved in the project and we finished developing the software, which is called Blur. Meanwhile, we could hear industry rumblings about public sharing of clinical trial documents becoming a new area of interest. It seemed a logical – although not an easy next step – to expand the Blur software to meet these new needs. I found myself at every possible industry conference trying to learn about clinical trial data and document transparency, and I developed an understanding of the responsible clinical trial data sharing that is such an important consideration in the writing of clinical documents, for example, CSRs. I heard that EMA was setting up a TAG, and I applied for a role. I was pleasantly surprised when I was selected, and so here I am on EMA's TAG.

Sam: The TAG includes members from academia, industry, patients, and healthcare professionals. It's interesting that there are no medical writers on the TAG because we are largely responsible for making decisions about and anonymising texts within CSRs (and clinical summary documents) that may compromise the privacy of patients when those CSRs are made publicly available. We currently do this in a qualitative way, by proactively anonymising the text in our CSRs to protect individuals, whilst trying to maximise data utility. We know that our statistician colleagues are learning about quantitative ways in which privacy can be protected. I believe that one such method is for statistical experts to develop structured statistical methodologies and that these may lessen the burden of medical writers … eventually. What are structured statistical methodologies, and how might they help us in our work?

Cathal: So, I need to give you some background here before I get to the nub of your question. When anonymising documents, it can be difficult to know which data might be "identifying" and which data might not. For example, if you have 100 participants from the USA and only one from Ireland, it is fairly obvious that to protect the participant from Ireland you would need to redact "Ireland", but you could retain "USA" in your document without compromising the USA-based participants. But what about if you have 20 people from Ireland, and only five of them are female? Once you start including other identifiers, such as people's race, age, or ethnicity, it can become quite easy to make an individual stand out as highly identifiable. This is where things get complicated, and present challenges for medical writers. So, statisticians use quantitative risk to establish which rules should be applied to identifiers (age, gender, race, country, etc.) to protect patient privacy whilst maintaining data utility. It is generally thought good practice to group some values together rather than to redact all information. Let's say that you have participants from Ireland, England, Spain, and France. There may be low numbers of participants from each country, but if you group them together and label them as

"Europe", it may make the participants less identifiable while maintaining some geographic information. We tend to apply similar grouping with numeric values such as age, height, and weight. Instead of completely redacting these values, we group them in numeric bands. These are very simple examples of how we apply structured statistical methodologies to statistically anonymise data. There are more complex methodologies for more complex situations.

Sam: So now we can see how our statistician colleagues can do some of the work for us by anonymising datasets in this way. I imagine it can get very complicated to anonymise clinical trial datasets in this quantitative way. There is the risk of reverse-engineering data to identify an individual if data anonymisation is not done properly. How do you address that aspect when developing structured statistical methodologies?

Cathal: We address this by considering a measure called "k-anonymity". In short, k-anonymity is how many people share the same characteristics. For example, if you have a small group of three people, all from the same country, who are the same, race, gender, ethnicity, and of similar height and weight, then you have a 1 in 3 chance of "guessing" which person is which. So you take the number 1 and divide it by 3. This gives you a quantitative risk score of 0.33. The recommended threshold for public sharing is 0.09. What this number actually means is that each participant must share the same characteristics as at least 10 other people in the population. Note that I said population and not clinical trial. The population refers to the number of participants that you are using as your total for sharing.

Determining your population is often the first part of calculating your risk. The larger the population, the more likely it is that patients are going to share similar characteristics to other patients. There are several possibilities when it comes to determining an appropriate population for calculating the quantitative risk. Appropriate populations include:

Document Population

"Document population" indicates that you are only going to use the patients that appear in your document as the population for your risk calculations. This can be tough to achieve. The documents being shared will often concentrate on a subset of the patients that were in a clinical trial. This can mean that you have a very small population. When you have a very small population, then patients can be unique with just a few identifiers.

Clinical Trial Population

"Clinical trial population" is all the patients that took part in the trial, upon which the documents being shared are based. As mentioned earlier, not all patients in a clinical trial are discussed in the document. When we use the clinical trial population, we can use a larger population to calculate the risk of reidentification of patients. The larger the population the more likely that patients will share similar characteristics.

All Similar Sponsor Trial Population

If you wish to calculate the risk of reidentification, you could use an even larger population. Sponsors tend to carry out more than a single trial within their therapeutic area. If you group the data about these patients from all these trials together, you end up with a much larger population. As pointed out earlier, larger populations mean it is more likely that patients will share similar characteristics and reduce the risk of reidentification.

All Similar Trial Population

This is almost exactly like the "all similar sponsor trial population" except that you calculate risk based on trials from other sponsors as well as your own. In practice, this can be hard to do as you normally do not have access to the data from other sponsors.

Therapeutic Area Population

This is usually the largest population considered when calculating the risk of reidentification. This would be everyone in a geographic area within a therapeutic area. For example, all diabetics in the USA. As you can imagine, this would be a huge population. The difficulty is that I am unlikely to have access to this individual information for every single diabetic in the USA. However, there are documented techniques for population estimators.

Which population to choose often comes down to a sponsor perception as to what an attacker is likely to know. Will they know the name of the trial that the person they are attempting to reidentify took part in?

Sam: It is a relief for medical writers to know that statistical colleagues are right behind us in protecting clinical trial data within datasets. This will eventually lessen our burden and will mean that we will have less proactive anonymisation and less redaction to do in our clinical reports. This is, however, clearly an industry "work in progress", and the fact that EMA has set up the TAG for this purpose shows its importance. In your opinion, how long until this becomes the normal way of working i.e. statisticians routinely applying structured statistical methodologies to clinical trial datasets before we medical writers start to report the data?

Cathal: I think that we are rapidly changing our current procedures to prepare for public sharing. However, people are nervous when we have different agencies with different criteria that they need to meet. Ideally, we would see all agencies aligning their policies so we have one harmonised way of publicly sharing documents. Unfortunately, we are not there yet. Right now, Health Canada is the main driving force, and people are adapting to their public sharing policy, which is well-aligned with EMA's Policy 0070, although there are still some slight differences.

References

1. EMA Technical Anonymisation Group web page: https://www.ema.europa.eu/en/human-regulatory/marketing-authorisation/clinical-data-publication/support-industry/external- guidance-implementation-european-medicines-agency-policy-publication-clinical-data
2. EMA Policy 0070 and implementation guidance web page: https://www.ema.europa.eu/en/human-regulatory/marketing-authorisation/clinical-data-publication/ technical-anonymisation-group
3. CDISC Standard, STDM: https://www.cdisc.org/standards/foundational/sdtm
4. CDISC Standard ADaM: https://www.cdisc.org/standards/foundational/adam
5. CDISC home page: https://www.cdisc.org/new-to-cdisc

A role for medical writers in overcoming commonly held misconceptions around FAIR data

Erik Schultes
International Science Coordinator
GO FAIR International Support and Coordination Office

Correspondence to:
Erik Schultes
Poortgebouw N-01
Rijnsburgerweg 10
2333 AA Leiden
The Netherlands
erik.schultes@go-fair.org

Abstract
More and more, computers must participate with physicians and patients as trusted partners in assessing medical options and tracking outcomes. But before the computer can become a routine medical assistant, data and computational services must become Findable, Accessible, Interoperable and Reusable (FAIR) for machines. In this article, the fundamentals of FAIR data and some technology trends are described, with clarification of commonly held and often repeated misconceptions about FAIR. As FAIR was conceived primarily in the life sciences, medical writers are uniquely positioned to help counter these misconceptions.

In March 2016, a little more than 2 years after the initial Lorentz Centre Workshop that launched the now well-known acronym,[1] a brief commentary appeared in the journal *Scientific Data* that quietly enunciated the FAIR (Findable, Accessible, Interoperable and Reusable) Guiding Principles for data stewardship.[2] The commentary gave a vision for a world where data and services could automatically interoperate and framed some of the current barriers to getting there. The FAIR Principles themselves were composed of 15 one-liners, were presented as a nondescript figure item (Box 2 in the commentary), and were accompanied by no in-depth discussion about how they should actually be implemented. Despite this innocuous beginning, the FAIR Principles would go on to take the world by storm.

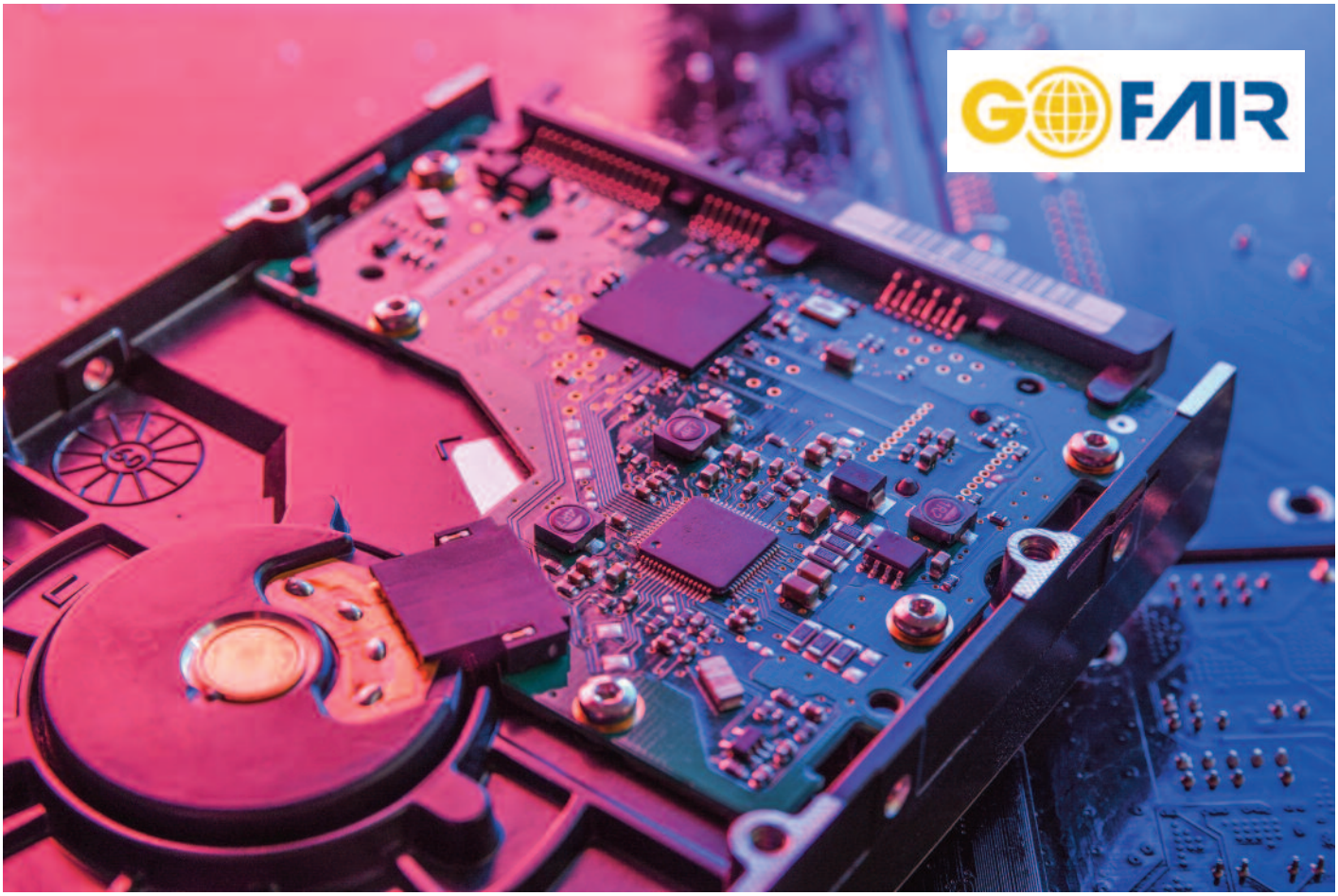
The FAIR Principles make digital data and services findable, accessible, interoperable, and reusable not only to human users (e.g., doctors, patients, researchers who may be sitting behind computers or using smart phones) but also, in more and more automated ways, by computers themselves. Despite the egalitarian ring to the FAIR acronym, the FAIR Principles are actually concerned with only the technical issues regarding data handling, although data that are FAIR for machines are arguably fairer for humans.

The FAIR Principles emerged in response to the key conundrum posed by big data: although combining data from multiple sources into ever-larger collections creates immense value and opportunity, the rapid growth rates of data production mean than that such data collections cannot be built manually. Often the data are hidden in plain sight, but even when they are known, the human intervention required to process them makes their reuse impractical.[3] Without FAIR, future investments in big data, data analytics, and AI will bring only diminishing returns.[4] Under the growing weight of this conundrum, publicly declared endorsements of and commitments to the FAIR Principles came suddenly and decisively from across the stakeholder community. Among the earliest to commit were high-level organisations that wanted to harness big data analytics for high-value applications. These included powerful economic development alliances like the G20,[5] the G7,[6] and the World Economic Forum.[7] Almost as soon as the FAIR Principles were published, the European Commission announced its ambition to build the "European Open Science Cloud", a revolutionary data infrastructure based on the FAIR Principles that would safeguard member states' research data, particularly data produced with public funds (totalling €300 billion annually).[8,9] Similar large-scale public data infrastructure based on the FAIR Principles has also been launched in other countries and regions including the US, China, and Africa.[10] Now, 4 years after their publication, we begin to see serious discussions on the implementation choices and challenges around the FAIR Principles from those stakeholders in the trenches who are actually "doing" FAIR data. In addition to a growing list of publications,[11] research projects,[12] and initiatives,[13] private industry[14] is also beginning to offer solutions to help make FAIR data and services more practical.

With this explosive development of FAIR, a number of misconceptions about FAIR have taken root throughout the stakeholder community, including among policy makers and self-proclaimed practitioners of FAIR.[15] Four examples are given below, together with a few tips and tricks to keep in mind when writing about FAIR. Because FAIR originated largely in the life sciences and has been driven oftentimes by applications in the biomedical space, medical writers are uniquely positioned to help mitigate, if not actually rectify, these common misconceptions about what FAIR is, what it is not, and its relevance to various stakeholders.

First, although the ultimate purpose of FAIR is to help people (e.g., patients, doctors, epidemiologists, insurance companies) to solve complex problems and make better-informed health and medical decisions, the FAIR principles are directed at machine agents. "FAIR for machines" is therefore a redundant phrase.

Second, FAIR is not a data format, computer protocol, or standard. Rather, it is a set of principles which can be applied to data formats, computer protocols, and standards to ensure they are machine-actionable. There tends to be broad and general agreement around the FAIR Principles (like there is for world peace), but finding widespread agreement on implementation choices for FAIR can be a challenge. Among other mechanisms, the publicly funded GO FAIR International Support and Coordination Office (go-fair.org), a joint project launched by ministries in the Netherlands, Germany, and France, has as its mandate the acceleration of bottom-up convergence on standards and technologies to achieve broadly



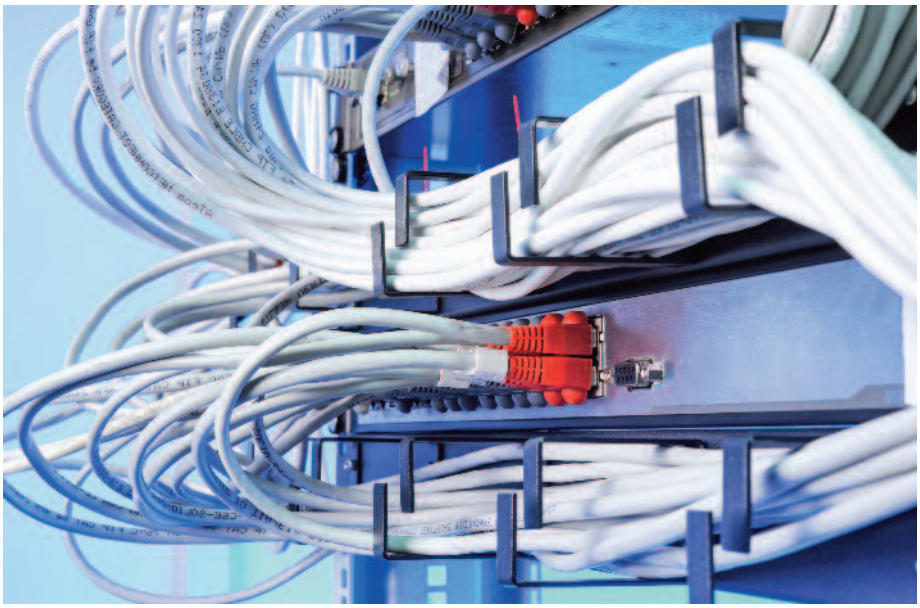
accepted FAIR implementation.

Third, and highly relevant for medical and pharmaceutical data, FAIR is often confounded with "open source", "open access", and "free". The principles under the "A" in FAIR stipulate only that the conditions to access data are so clearly spelled out that even a machine will know exactly what actions it must take before it can perform operations like data query or advanced analytics.[16] Hence, sensitive data like patient records may be rendered perfectly FAIR, but because of GDPR and other privacy restrictions, will never be made open. In other cases, like FAIR data created within private companies, part of the access procedure could be to accept a restrictive licence and pay a fee. The data in these cases are not open or free but could (and should) be made FAIR. Separating FAIR and open makes it possible for data owners to preserve control over their data while at the same time allowing these data to interoperate with other resources if and when needed. New approaches allow for fine-grained access control, allowing certified algorithms to access some data elements (e.g., patient blood pressure measurements) but not others (e.g., patient personal identification data). These types of applications open the door to so-called "distributed learning applications", also known as FAIR Data Trains.[17] In this approach, certified algorithms (Trains) are dispatched to the data (FAIR Data Stations) where access is given to whatever data are allowed in that particular case. FAIR Data Trains turn the traditional idea of data sharing upside down, giving control of data access to data owners (perhaps even patients themselves) who in turn grant permission to algorithms that visit the data. "Data visiting" rather than "data sharing" will open up the tremendous knowledge stored in patient records to researchers, while privacy and personal information remain protected.

Fourth, FAIR often raises concerns over data quality and trustworthiness: Might data (e.g., personal health data from wearable devices) be published without checks like peer review? How can we trust FAIR data? Oftentimes, researchers and policy makers propose to expand FAIR with additional principles like "Ethics" and "Responsibility" (FAIR-ER data), but by design explicit references to data quality were not included in the FAIR Principles. Data quality is context-dependent, where data of high quality in one case might be viewed as low quality in another case. The FAIR Principles prompt the creation and use of "rich" machine-readable provenance metadata (Principle R1.2), which is often missing in conventional datasets. Provenance metadata refers to any and all relevant information about the creation of the data: when and where the data were produced, and by whom; the source of funding for the data acquisition; and the methods and instruments used to collect and analyse the data. These and many other elements belong in properly constructed FAIR data. Such metadata will, by inference (either human or machine), only improve the trustworthiness of the data and ensure more accurate assessment of data quality.

After 4 years and more than 2600 citations, the FAIR Principles have already made an impact on data-intensive disciplines, especially in health, medicine, and biomedical sciences. Undoubtably, there will much more to say about FAIR in medicine in the coming years. Of all the many initiatives around FAIR, the most recent and most urgent is driven by the COVID-19 pandemic. A newly minted GO FAIR Implementation Network called the Virus Outbreak Data Network (VODAN),[18] composed of public and private regional authorities, is now attempting to install an international network of FAIR Data

Stations across the world and make COVID-19 electronic case report forms available for automated global data visiting. Driven by the urgency of the disease outbreak, FAIR is being put to the test on a scale never attempted before. Accurately communicating the aims and methods of FAIR in the coming years will continue to be a challenge, but the resources needed by journalists and medical writers to decipher the FAIR Principles are there, if you know where to look.[11,19,20] One day, even these resources themselves might be made FAIR.



Conflicts of interest

The author declares no conflicts of interest.

References

1. Dutch Techcentre for Life Sciences. Jointly designing a Data FAIRPORT. 2014 Jan 20. Available from: https://www.dtls.nl/2014/01/20/jointly-designing-data-fairport/
2. Wilkinson MD, Dumontier M, Aalbersberg I, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 2016;3(160018).
3. Press G. Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says. 2016. Available from: https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/#52a441026f63
4. PwC EU Services. Cost of not having FAIR research data. Brussels: Directorate-General for Research and Innovation, 2019.
5. European Commission. Article 12, G20 Leaders' Communique Hangzhou Summit. 2016. Available from: https://ec.europa.eu/commission/presscorner/detail/en/STATEMENT_16_2967
6. G7 Science and Tsukuba Communiqué. G7 Science and Technology Ministers' Meeting. 2016 May 15-17. Tsukuba, Ibaraki.
7. World Economic Forum. Value in Healthcare: Mobilizing cooperation for health system transformation. Geneva, Switzerland: World Economic Forum; 2018.
8. European Commission. Communication from the Commission to the European Parliament, the Council, The European Economic and Social Committee and the Committee of the Regions. Brussels, Belgium: European Commission; 2016.
9. Mons B. Invest 5% of research funds in ensuring data are reusable. Nature. 2020;578(7796).
10. Meerman B. OPEN SCIENCE CLOUD INITIATIVES: Region = FAR EAST. 2019. Available from: https://www.eoscsecretariat.eu/eosc-liaison-platform/post/open-science-cloud-initiatives-region-far-east
11. Mons B, Schultes E, Liu F, Jacobsen A. The FAIR Principles: first generation implementation choices and challenges. Data Intelligence 2020;2(2020):1-9. Available from: http://www.data-intelligence-journal.org
12. European Union Horizon 2020. FAIRsFAIR "Fostering FAIR Data Practices In Europe". European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 Grant agreement 831558. Brussels, Belgium: European Commission; 2020.
13. GO FAIR. FAIR Made Easy for Funders. Available from: https://www.go-fair.org/today/FAIR-funder
14. Phortos Consultants. The FAIR Service Provider Consortium (FSPC). Available from: http://www.phortosconsultants.com/Consortium
15. Mons B, Neylon C, Velterop J, et al. Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. Inf Serv Use. 2017;37(1):49-56.
16. Jacobsen A, de Miranda Azevedo R, Juty N, et al. FAIR Principles: interpretations and implementation considerations. Data Intelligence. 2020;2(1):10-29.
17. Dutch Techcentre for Life Sciences. Personal Health Train. 2016. Available from: https://vimeo.com/143245835
18. GO FAIR. Virus Outbreak Data Network (VODAN). 2020 [cited 2020 Apr 21]. Available from: https://www.go-fair.org/implementation-networks/overview/vodan/
19. Kryger Hansen K, Buss M, Sztuk Haahr L; FAIR project. A FAIRy tale. Available from: https://vidensportal.deic.dk/sites/default/files/uploads/A%20FAIRy%20tale%20book%20digital_ny.pdf
20. Go FAIR. FAIR Principles. Available from: https://www.go-fair.org/fair-principles/

Author information

Erik Schultes is an evolutionary biologist with a long-standing interest in data-intensive research questions. As International Science Coordinator at GO FAIR, Erik has worked with a diverse community of stakeholders to develop FAIR data and services, both within and outside the life sciences.
https://orcid.org/0000-0001-8888-635X.

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
- Contributions
- Editorial Board

Website
- Contributions
- Web Team

Freelance Business Group

Social Media Team

Training
- Leading workshops
- Professional Development Programme Committee
- Webinar contributions
- Webinar Team

Special Interest Groups
- Pharmacovigilance
- Regulatory Disclosure
- Medical Devices
- Medical Communications
- Veterinary Medical Writing
- Sustainability

Ambassador programme

Executive Committee
- President
- Vice President
- Journal Editor
- Public Relations Chair
- Conference Chair
- Honorary Secretary
- Professional Development Programme Committee Chair
- Treasurer

TO FIND OUT MORE

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

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

MyData: Applying human-centric principles to health data

Authors:

Mad P. Ball PhD [1,2,3*]
Sherry Chung MSc [1,4,5*]
Isabelle de Zegher MD, MSc [1,6*]
Christoph Fabianek PhD, MSc, MBA [1,7,8*]
Casandra Grundstrom MSc, BTech [1,9,10*]
Paul Knowles BSc[1,11*]
Viivi Lähteenoja MA [1,3,12*]
Joss Langford BSc, FRSA, MIET [1,13,14,15*]
Fredrik Lindén [1,16,17,18*]
Maritta Perälä-Heape PhD [1,9,10*]
Moira Schieke MD [1,19,20*]
Berrin Serdar MD, PhD [1,21,22*]
Dixon Siu MSc, BEng [1,23,24*]
Chi-Chuan Wang PhD, MHPA [1,25*]
**All authors contributed to this article equally*

Affiliations:

1 MyData Health Thematic Group
2 Open Humans Foundation, United States
3 MyData Global ry, Helsinki, Finland
4 Numbers Co., Ltd, Taiwan
5 MyData Taiwan Hub
6 b!loba, Tervuren, Belgium
7 OwnYourData, Austria
8 MyData Austria Hub
9 University of Oulu, Oulu, Finland
10 MyData Oulu Hub
11 The Human Colossus Foundation, Geneva, Switzerland
12 University of Helsinki, Helsinki, Finland
13 Activinsights Ltd, Kimbolton, UK
14 University of Exeter, Exeter, UK
15 Coelition, Exeter, UK
16 Grandezza Konsult AB, Stockholm, Sweden
17 MyData Sweden Hub Non-profit Association, Lidingö, Sweden
18 Grapevine World LLC, Vienna, Austria
19 University of Wisconsin, Madison, Wisconsin, United States
20 Cubismi Inc, Madison, Wisconsin, United States
21 Cloud Privacy Labs LLC, Greenwood Village, Colorado, United States
22 University of Colorado Denver, Aurora, Colorado, United States
23 Personium Project, Tokyo, Japan
24 MyData Japan Hub
25 National Taiwan University, Taipei, Taiwan

Correspondence to:

Casandra Grundstrom
MyData Health Thematic Group
MyData Global ry
Maria 01
Lapinlahdenkatu 16
00180 Helsinki, Finland
health@mydata.org

Abstract

Modern data legislation increasingly empowers citizens, and therefore patients, with rights to access and control their health data. The mechanisms needed to exercise modern data rights are currently underdeveloped and underserving individuals and societies. MyData is the human-centric approach to shift the power of personal data more equitably into the hands of individuals as part of a fair data economy. In this article, we present different scenarios that apply the MyData principles for human-centric control of health data. These scenarios demonstrate the potential of the human-centric approach for turning data rights into truly actionable points for policymakers, healthcare stakeholders, and medical communicators.

The emergence of the health data economy

The ability to digitalise health records has not only revolutionised the practices of clinicians and healthcare organisations, but it has also started a social change that is radically reforming the relationship between the individual and their health data.[1]

> "Numbers can't speak for themselves, and data sets – no matter their scale – are still objects of human design." Kate Crawford

Integrated, high-quality health data is a potential treasure-trove for healthcare. Diverse actors seek to leverage the digitalisation of healthcare to develop data-driven benefits – a phenomenon that is emerging as the "data economy" (see Figure 1).[2] Ideally, in the health data economy, providers can improve efficiency from seamless, centralised access to longitudinal records all while individuals immediately access and control their health data. However, when health data is framed exclusively in terms of the data economy, it runs the risk of reducing individuals to mere

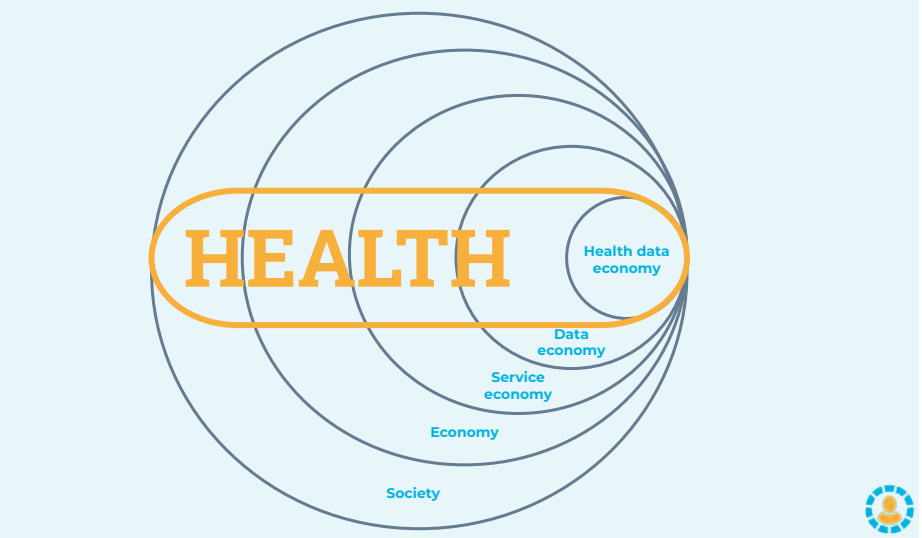


Figure 1. The context of health data in the economy
The health sector as a whole reaches across all aspects of modern economies, while the data economy has emerged as a driving force of innovation behind the service economy. The health data economy has the potential to play an increasingly significant role in both healthcare and the data economy.

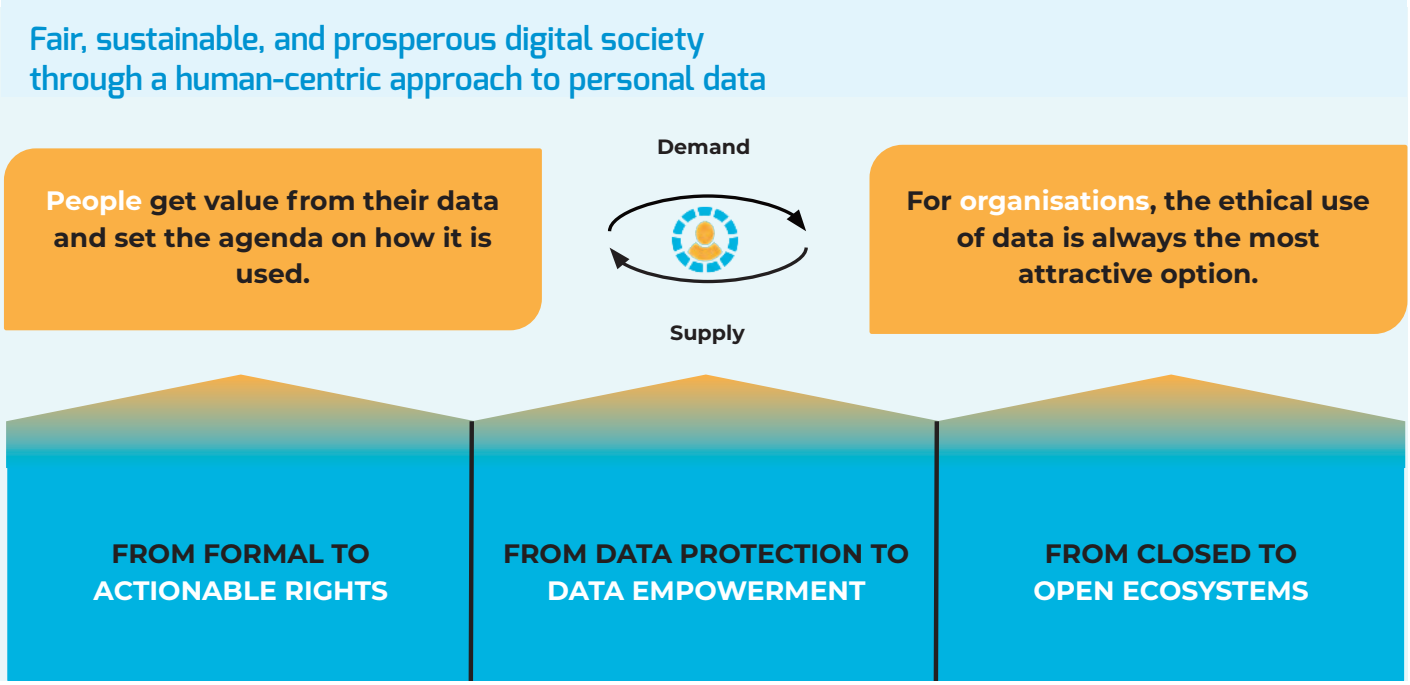


Figure 2. Goals of the MyData approach
To build a fair, sustainable, and prosperous digital society, three shifts are necessary, as represented by the boxes at the bottom of the figure.

market participants. Consideration should be given to the wider societal context beyond the data economy to support human agency and ultimately advocate for the role and rights of citizens.

MyData

MyData is a human-centric approach to personal data.[3] This human-centric paradigm was introduced to support the realisation of "a fair, sustainable, and prosperous digital society, where the sharing of personal data is based on trust as well as a balanced and fair relationship between individuals and organisations" (see Figure 2).[4] MyData Global is a non-profit organisation catalysing the adoption of this approach, aiming to empower people by improving their right to self-determination in all aspects of personal data.[5] Driven from the ground up, the MyData community published a declaration outlining the shifts needed to reach the human-centric paradigm and the principles upon which to build it (see Figure 3).[6] The European Strategy for Data, a recent communication released by the European Commission, endorsed the MyData approach as a promising initiative to actualise human-centricity.[7]

The principles

Multiple, interrelated issues exist in the current *modus operandi* of the health and wellness ecosystem, many of which can be addressed through the MyData principles (see Figure 3). The fractured state of the contemporary health data landscape is one of the fundamentally problematic aspects. The MyData approach to remedy this situation is to adopt the principle of the individual as the point of integration; that is, the individual acts as the central contact point to access and control their data. This approach simplifies the challenges around consent, enables human-centric governance, and creates opportunities for service innovation.

Healthcare organisations are facing risks associated with being data custodians, including the burden of security and the possibility of severe penalties for data breaches and privacy violations.[8] By adopting the principles of transparency and accountability, organisations can prepare for intended and unintended consequences from their use of health data in a manner that creates trust and mitigates risks. Making a conscious shift towards human-centric control of personal data, organisations in the health and wellness sector empower people to be healthier and flourish with the help of the data available about them.

Modern data protection and privacy legislation grants individuals the rights over their personal data. However, the tools, skills, and opportunities necessary to exercise these rights are often underdeveloped and overlooked. Adopting the principles of interoperability enables the sharing of health data between organisations and individuals, while portability allows individuals to access and use data about them. These two principles can help organisations to comply with existing and emerging legislation, and better serve their customers. In short, the MyData approach is an attractive option for organisations because it ensures individuals are empowered to control their data and promote its best use.[9]

In short, the MyData approach is an attractive option for organisations because it ensures individuals are empowered to control their data and promote its best use.

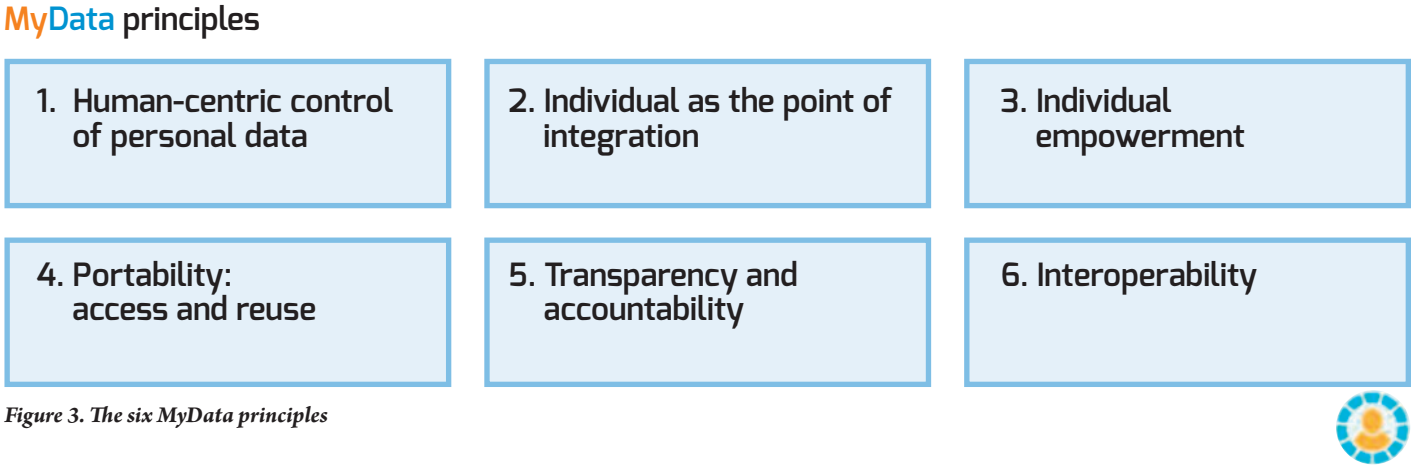


Figure 3. The six MyData principles

The power of applied MyData thinking

In this section, we demonstrate the power of the MyData approach by discussing four contemporary challenges in healthcare and presenting innovative solutions and initiatives that succeed in addressing these challenges through the adoption of MyData principles.

Health data access and reuse for patients with diabetes

Access to health data has proven beneficial for self-management of complex chronic diseases such as diabetes.[10] Diabetes management tools (e.g., insulin pumps, blood glucose meters, and continuous glucose monitors [CGMs]) provide a rich source of health data, yet, the potential value of this data is not maximised to meet all desired uses by patients. While device manufacturers provide patients with software tools for data use, MyData principles underlie advances in diabetes data reuse by patients for other purposes. In particular, three relevant principles apply: individuals as a point of integration, individuals as autonomous agents in the reuse of their data, and the right of individuals to port their data to their own devices and other services. These principles have come together to give patients the ability to create and use data for purposes beyond those provided by device manufacturers.

The huge variety of tools, grassroots, and non-profit efforts for repurposing diabetes data highlight the progress in this domain. The patient-initiated Nightscout Project is an open-source software project for CGM data portability, enabling coordinated care and data access for caretakers as well as re-display of data on different devices. Patients have also developed open-source tools for "looping" (automated insulin delivery personalised from CGM data). Remarkably, due in part to government regulation, these efforts made automated insulin available to patients before commercial products were available. Data integration across devices is also enabled with open-source software developed by Tidepool, and the non-profit organisation has launched a project to develop an FDA-regulated version of Loop software.[11] Patients have also been able to donate data to research via a centralised database managed by Tidepool as well as similar patient-led research projects.[12,13] The ability of patients to access and reuse their data as advocated by MyData principles has enabled a remarkable set of innovations beyond what would have otherwise existed.

Empowering individuals with control of their health data

Healthcare is challenging due to the sensitivity of health data, the imbalance of power in the patient-provider relationship, and healthcare providers' historical status as controllers of data. Empowering individuals with control over their health data is a key MyData principle that addresses this problem. When individuals are given control over their data, they are more willing to share their health information.[14] Certain types of health data are more sensitive, such as those related to mental health or substance use disorder, and require an individual to decide what can be shared, with whom, for what purpose, and for how long. Understanding individuals' perceptions regarding what specifically matters for their privacy is essential.

A form of control is enacted through informed consent, which has been at the centre of human subject research since the Declaration of Helsinki in 1964.[15] However, consent terms and privacy policies often include lengthy legal text that is not conducive to engaging individuals in informed decision making. Lack of meaningful choices can cause an individual to deny consent or avoid medical services out of fears of discrimination or abuse of power. For medical providers, timely access to the right data at the right time is paramount, and the scope of access is dependent on the context. For example, data accessible to a physician during a regular visit can differ from data available during substance use disorder treatment. Emerging platforms empower individuals by integrating their privacy concerns into data flow based on context while enabling centralised control over health data.[16,17]

To provide individuals with a central point of contact to manage their data, open-source solutions for personal data stores are becoming increasingly relevant.[18,19] Innovations which empower individuals with control over their health data are crucial for balancing power in the data economy.

Data interoperability

Data interoperability standardises interfaces, data syntax, and the semantics of the underlying data, going beyond technical integration to ensure that data has the same meaning wherever it is used. It provides the best opportunity for the use of health data and appropriate standards across different types of systems as advocated for by the MyData principles. The ability to integrate universally at the data level is known as semantic harmonisation, and this provides full data portability. This is desirable because storing data together with the tools necessary for interpretation makes it easier for a wide range of services to access the information. A good example of a storage tool that achieves this is a Semantic Container.[20] Tightly controlled data formats also improve interoperability capabilities from new

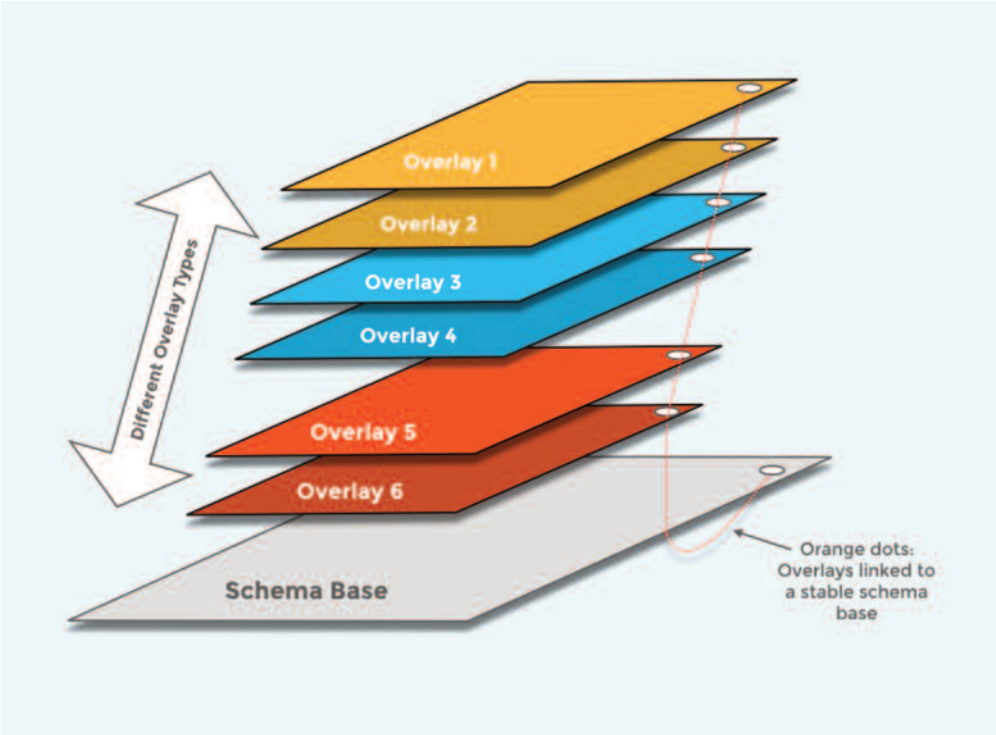


Figure 4. Overlays capture architecture can enable object interoperability within an industry sector
This architecture presents a schema base and interoperable overlays. Different overlays represent the specific needs of a particular application or organisation with, for example, dedicated formats, character sets, or information overlays. Multiple parties can interact with and contribute to the overlays without having to change the schema base definition.

applications such as Internet of Things devices.[21]

A universal approach to a unified data language would provide data subjects with the opportunity to request enriched data, in a standard format for self-governed usage, from trusted brokers such as national governments, insurers, or healthcare providers. A layered capture architecture (see Figure 4) can facilitate data harmonisation beyond the confines of a single organisation by enabling a number of parties to contribute to multi-source data pools. This level of object interoperability and subsequent language unification will provide increasing efficiency and effectiveness within the healthcare sector.

A core component to facilitate a unified data language is object interoperability using a schema with a layered architecture.[22] By introducing layered task-orientated objects (or "overlays") multiple parties can interact with the same base object by simply creating their own set of linked overlays to fulfil the requirements of their particular use case. With schema base definitions a stable foundation is established to capture health data in a standardised way.

Any specific machine-readable task, rule or definition relating to the semantics of the source data can be encapsulated into a data capture object. To maintain the reusability of data capture objects, each item should remain accessible to all developers interacting with the network. Although there must always be an auditable trace back to the public identifier of the original issuer of a data capture object, that actor never needs to govern who has access to those published items. All actors will benefit from the availability of data capture objects in open form which, alongside a human-centric design ethos, helps to drive interoperability through standardisation.

Establishing trust between stakeholders for health data use

Traditionally, trust between patients and their caregivers is a necessary part of medicine and healing. However, this trust has been abused in some cases, and as a result, new data policies are in place to protect patients.[23] With an increase of health data, the relationship between clinicians, healthcare organisations, and patients is evolving and questions of how trust can be achieved are necessary as part of a fair health data economy.

As the generation of health data often involves multiple parties, it can become difficult to determine who governs data, which impedes data use. One way to rebuild trust between healthcare systems, data vendors, and individuals is for stakeholders to embrace the MyData principle of transparency and accountability. Although informed consent provides specific terms and conditions for data use, there are often grey areas where data can potentially be used beyond the scope of the agreement. Traceable health information services allow individuals to have full access and control of health data about themselves through a transparent data management process.[24] The accountability of data flow reduces individuals' worries about data misuse.

Because health data can be copied with ease, special measures are necessary to ensure terms of use. Digital watermarking provides an elegant way to foster transparency and accountability in data transfers, increasing trust between

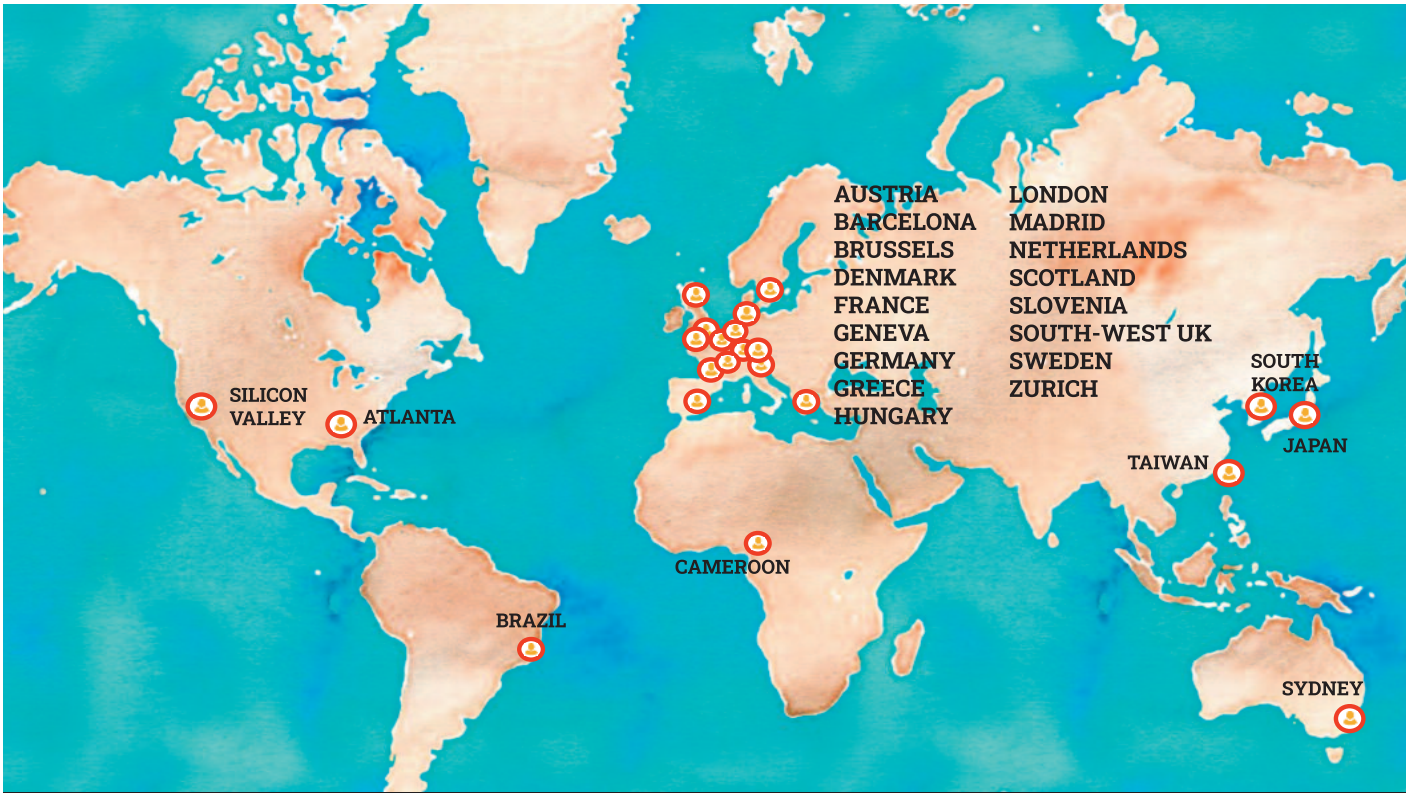


Figure 5. Locations of MyData hubs

individuals and healthcare stakeholders. The individual benefits by proving lawful access to the data all while building trust through transparency mechanisms. In practice, Semantic Containers provide an open-source implementation for digital watermarking.[20] Ultimately, traceability of health data invokes the necessary level of trust to encourage data sharing on a global scale for societal benefit.

The imminent role of MyData in the data economy

MyData for healthcare stakeholders

The four areas of application outlined above clearly reveal the interplay between individuals and organisations of the healthcare sector. While we have shown that technologies are emerging to support these dynamic relationships, in each case, it is human-centric design thinking that achieves the balance required for a sustainable model. The initiatives, companies, and services presented here expose a world where new tools for exchanging health data are in the process of being adopted. These tools are starting to open up the large health data repositories and monopolies by making the individual the point of integration. From the point of view of healthcare organisations, the health economic arguments for these human-centric approaches are undeniable–they reduce regulatory burden and risk, provide access to new sources of data, and build equity with individuals.

MyData for policymakers

Data protection regulations in many countries increasingly provide rights to individuals; however, this should not be seen as the sole driver for providing increased access to and control over health data. Considerable benefits also arise from data sharing, portability, and use. The European Commission's data strategy supports a common health data space as crucial for "advances in preventing, detecting, and curing diseases as well as for informed, evidence-based decisions to improve the accessibility, effectiveness, and sustainability of the healthcare systems".[7] Adopting the MyData principles from the top down would continue to support the individual as the point of integration for developing this shared health data space.

MyData for medical communicators

The approach advocated by MyData is to focus on human factors as central to the activities of the health data economy. Since medical communicators traditionally gather and disseminate health information from trustworthy sources, they are in an excellent position to convey the importance of best data practices in the medical field with confidence and sensitivity. Understanding the expectations, opportunities, and risks that individuals face is critical in the creation of informative materials for public dissemination. Medical writers should recognise the data rights of individuals and use MyData principles as a resource to empower human-centric communications.

Conclusion

The role of health data in the data economy as the source of predictive, preventive, personalised, and participatory power is an emerging phenomenon. Leveraging the abundance of health data as a source of value for relevant stakeholders can only be materialised through deliberate actions by policymakers, support from medical writers for disseminating best data practices, and innovative data management models embraced by organisations.[25] All healthcare actors have a vested interest in empowering individuals to improve health and wellness while advancing the health data economy. This should be approached

in a fair, transparent, and sustainable manner as advocated for by the MyData principles.

Acknowledgements

CCW was partially funded by a research grant from the Ministry of Science and Technology in Taiwan (MOST-107–2636-B-002–002). The funders had no role in the manuscript preparation or publication decisions.

Disclaimers

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

Conflicts of interest

The authors declare that they are all active members of MyData Global ry. MPB is also an elected member of the MyData Global ry board. VL is also actively employed by MyData Global ry.

References

1. Bhatt A. Evolution of clinical research: a history before and beyond James Lind. Perspectives in clinical research. 2010;1(1):6–10.
2. Tang C, Plasek JM, Bates DW. Rethinking data sharing at the dawn of a health data economy: A viewpoint. J Med Internet Res. 2018;20(11):e11519.
3. Finnish Ministry of Transport and Communications. MyData – A Nordic model for human-centered personal data management and processing. 2015 [cited 2020 Feb 29]. Available from: http://julkaisut.valtioneuvosto.fi/handle/10024/78439
4. MyData Global. 2020 [cited 2020 Feb 29]. Available from: https://mydata.org/
5. MyData Bylaws. MyData Global. 2020 [cited 2020 Mar 8]. Available from: https://mydata.org/bylaws/
6. MyData Declaration. MyData Global. 2017 [cited 2020 Feb 29]. Available from: mydata.org/declaration
7. European Commission. A European strategy for data. Brussels; 2020 [cited 2020 Feb 29]. Available from: https://ec.europa.eu/info/sites/info/files/communication-european-strategy-data-19feb2020_en.pdf
8. HIPAA Journal. Healthcare data breach report. 2020 [cited 2020 Feb 29]. Available from: https://www.hipaajournal.com/2019-healthcare-data-breach-report/
9. McCoy MS, Joffe S, Emanuel EJ. Sharing Patient Data Without Exploiting Patients. JAMA. 2020;323(6):505–6.
10. Jackson SL, DesRoches CM, Frosch DL, Peacock S, Oster NV, Elmore JG. Will use of patient portals help to educate and communicate with patients with diabetes? Patient education and counseling. 2018;101(5):956–9.
11. Tidepool. Loop. 2020 [cited 2020 Apr 25]. Available from: https://www.tidepool.org/loop
12. Tidepool. Big Data. 2020 [cited 2020 Feb 29]. Available from: https://www.tidepool.org/bigdata
13. Greshake Tzovaras B, Angrist M, Arvai K, Dulaney M, Estrada-Galiñanes V, Gunderson B, Head T, Lewis D, Nov O, Shaer O, Tzovara A, Bobe J, Ball MP. Open Humans: A platform for participant-centered research and personal data exploration. GigaScience. 2019;8(6):1–13.
14. Spencer K, Sanders C, Whitley EA, Lund D, Kaye J, Dixon WG. Patient perspectives on sharing anonymized personal health data using a digital system for dynamic consent and research feedback: A qualitative study. J Med Internet Res. 2016;18(4):e66.
15. World Medical Association (WMA). Declaration of Helsinki: ethical principles for medical research involving human subjects. 2018 [cited 2020 Feb 29]. Available from: https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/
16. ConsentGrid. 2020 [cited 2020 Feb 29]. Available from: https://consentgrid.com/
17. Kantara Initiative. 2020 [cited 2020 Feb 29]. Available from: https://kantarainitiative.org/
18. Personium. 2018 [cited 2020 Feb 29]. Available from: https://personium.io
19. OwnYourData. Data Vault. 2020 [cited 2020 Feb 29]. Available from: https://www.ownyourdata.eu
20 OwnYourData. Semantic Container for Data Mobility. 2020 [cited 2020 Feb 29]. Available from: https://www.ownyourdata.eu/en/semcon/
21. OASIS-Open. Classification of everyday living (COEL). 2019 [cited 2020 Feb 29]. Available from: https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=coel
22. Medium. Overlays Data Capture Architecture (ODCA): Providing a standardized global solution for data capture and exchange. 2019 [cited 2020 Mar 7]. Available from: https://medium.com/@paul.knowles_52509/overlays-data-capture-architecture-odca-providing-a-standardized-global-solution-for-data-caeb1679137a
23. Neame RL. Privacy protection in personal health information and shared care records. Journal of Innovation in Health Informatics. 2014;21(2):84–91.
24. Numbers. Bring Trust Into Data. 2020 [cited 2020 Feb 29]. Available from: https://numbersprotocol.github.io/
25. Kemppainen L, Koivumäki T, Pikkarainen M. Emerging revenue models for personal data platform operations: When individuals are in control of their data. J Bus Mod. 2018;6(3):79–105.

MyData Health thematic group information

MyData Global is a registered association whose mission is to advocate for a human-centric approach to personal data. With just over two dozen hubs currently established on five continents, MyData Global represents hundreds of individuals and dozens of organisations as members (see Figure 5). As part of the MyData Global organisation, the thematic group for health data is a diverse collection of like-minded individuals with a wide range of expertise and experience on the health-related aspects of personal data. Our interests and activities are inspired by health topics ranging from debating the status of genomic data, to knowledge creation on interoperability and data sharing. We advocate for the human-centric MyData principles and are an inclusive community. If you would like to learn more, or are interested in joining our community, you can find us on the online communication platform Slack for MyData[4] and on Twitter.

Thriving in the data economy as data-fluent PhD graduates

Menorca Chaturvedi
University Hospital Basel, Switzerland
University of Basel, Switzerland

Correspondence to:
Menorca Chaturvedi
menorca.chaturvedi@gmail.com

Abstract
We are living in an era of data deluge. With the amount of data generated increasing rapidly, organisations are in high need of individuals who are skilled at managing, analysing, and interpreting data. Data literacy is deemed as a crucial twenty-first century skill and one that will be required for a wide variety of roles. PhD graduates are a group of highly skilled professionals, many of whom have experience with handling different types of data during their research work and communicating the impact of their research to varying audiences. In this article, I look at how these professionals can add value to organisations, especially by combining their data analysis and communication skills, which can be applied across different roles in industry.

From smartwatches and fitness trackers alerting people to their health and sleep patterns, food technology apps working towards giving people an accurate picture of their personal nutrition, telemedicine helping people far and wide with precision medicine and assistive technology aiding the ageing population, applications leveraging big data are all around us and are only increasing every year. It had been estimated that by 2020, 1.7 MB of data would be created each second for every person on earth.[1] Hospitals had previously been predicted to generate approximately 665 TB of data by 2015 due to increase in medical images and electronic health records,[2] and the human genome is estimated to need 3 GB of storage space.

With some claims of data being the new oil or at least being a valuable currency,[3,4] it is safe to say that we are indeed living in a data economy.[5]

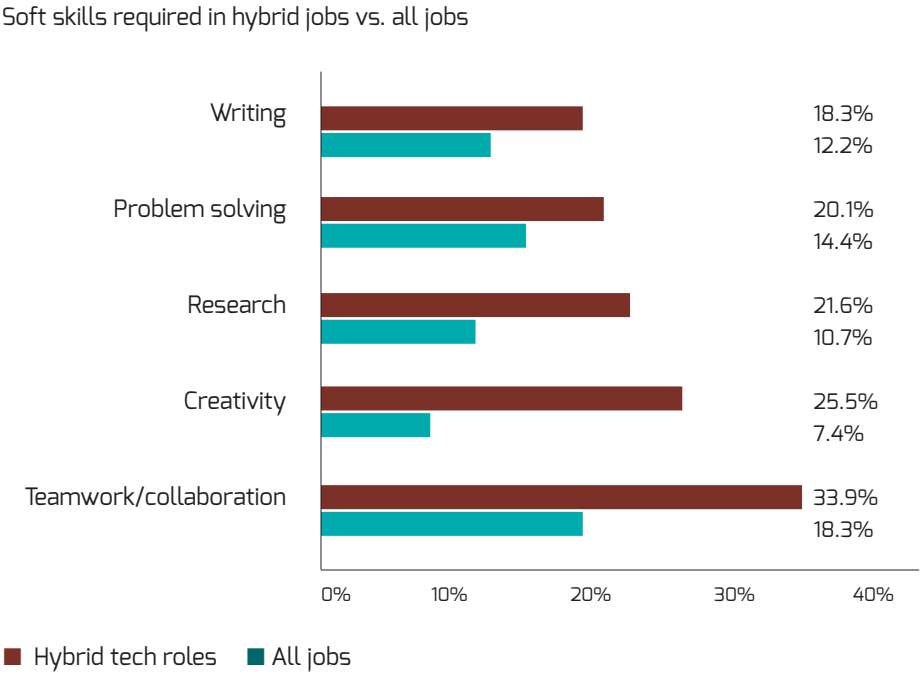


Figure 1. Soft skills required in hybrid jobs vs all jobs.
Source: Hybrid Job Economy Report 2018 by Burning Glass Technologies. Reprinted with permission.

Data literacy is recognised as a critical skill of the twenty-first century. In fact, poor data literacy was identified as a major challenge in the third annual Gartner Chief Data Officer Survey in 2017.[6] Gartner expects 80% of organisations to initiate competency development programmes in data literacy by 2020 to ensure data fluency across different roles, which would enable employees to manage, analyse, and apply data for value in context.

In the life sciences and healthcare sector, making sense of huge amounts of data and communicating its impact effectively is now more crucial than ever. The FDA has been working on the first-ever Digital Health Innovation Plan, Germany is making digital health applications reimbursable and the number of apps or devices being classified as medical devices is on the rise. So, it is likely that people with a strong knowledge and the ability to translate real-world data insights to different stakeholders would add value to organisations.

Hybrid jobs need data skills

An interesting term that could be used for such cross-functional roles is "hybrid jobs," a concept coined by Burning Glass Technologies in their 2018 report about the Hybrid Economy Job Market.[7] Hybrid jobs are said to be high-potential roles combining skills that were never needed together previously. For instance, this could be a combination of marketing and statistical analysis skills, or design and programming, just to mention two of their ideas. Despite being technology-driven, such roles may need more interpersonal skills like judgement, creativity, and collaboration. The report states, "Fully one-quarter of all occupations in the U.S. economy show strong signs of hybridisation, and they are almost universally the fastest-growing and highest-paying – and also the most resistant to automation. Some of these jobs are new, some are new versions of existing jobs, but all of them pose many different challenges for workers, students, employers, and educators."

The number of occupations with 10,000 or more job postings requesting creativity rose from 14 to 35 between 2012 and 2018. That includes roles like computer systems engineers, IT project managers, and programme managers.

The number of occupations requiring data science and analysis as skills and with 10,000 or more job postings also increased more than twofold between 2012 and 2018. While initially analytical skills were more common in technical requirements for jobs like for systems analysts, business analysts, database administrators, software developers, they started being mentioned for other roles like product managers, HR specialists, and even retail store managers in 2018.

One group of professionals rigorously trained in working on innovative ideas, handling different types of data and communicating with different audiences are PhDs. The number of PhD researchers globally has been growing continuously,[8,9] while the number of graduates pursuing tenure-track academic positions has declined.[10] The majority of the science and engineering PhDs use their technical, transferable, and soft skills to transition into different roles in industry. So, how effective is PhD training and what careers are these highly skilled professionals taking up in the data economy?

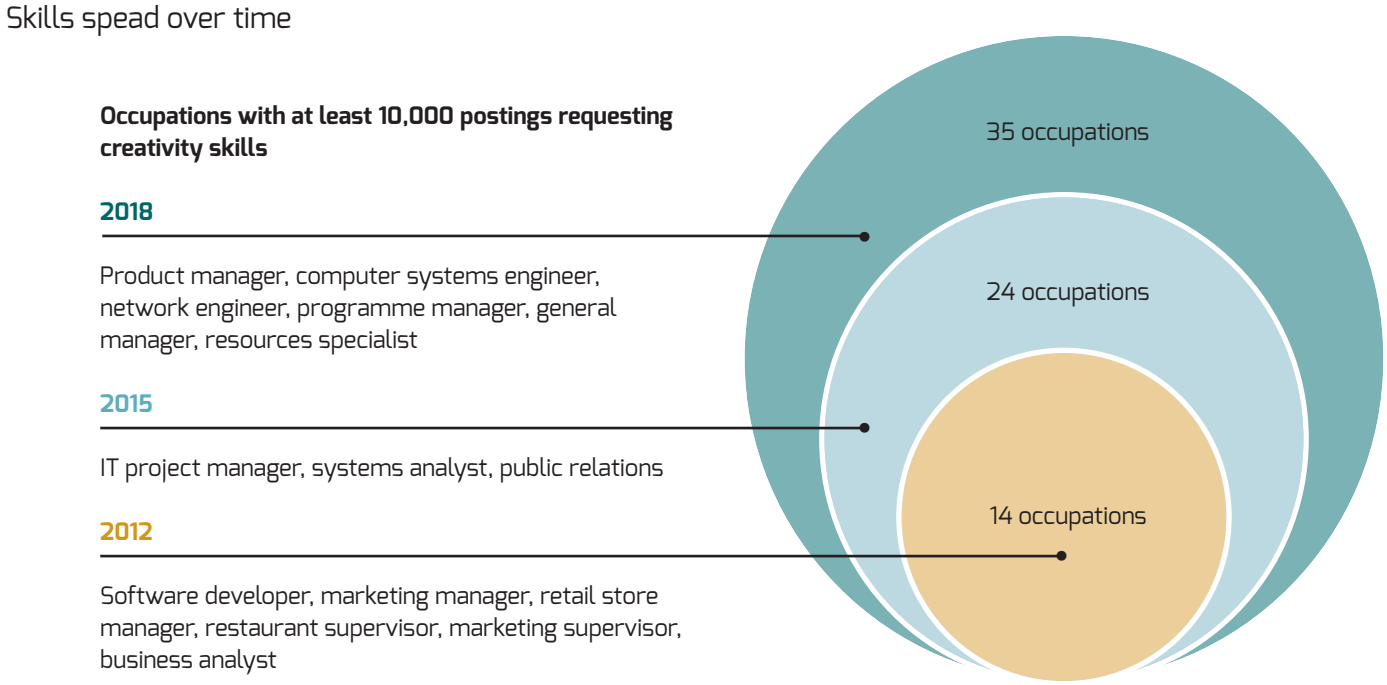


Figure 2. Occupations requesting creativity as a skill, in 2012, 2015, and 2018.

Source: Hybrid Job Economy Report 2018 by Burning Glass Technologies. Reprinted with permission.

The professional value of having a PhD

Investing 3 to 5 years on average in a PhD programme requires immense patience, grit, problem-solving skills, and, of course, the dedication and motivation to pursue knowledge while creating new knowledge. It teaches one to handle situations in the face of uncertainty and quickly change directions if projects don't go as planned. Researchers have to often participate in or lead collaborations, which could involve multiple centres, whether national or international. This, along with the fact that multidisciplinary projects are quite common, provides a good platform to interact with different stakeholders, handle expectations and manage projects while dealing with real-world data and its challenges.

One example of PhD programmes bringing various stakeholders together is the Marie Curie Innovative Training Network, which is an EU funded programme. The in3 project, funded by this initiative, focuses on reducing animal testing in the pharmaceutical industry. Fostering collaboration between 15 young researchers across Europe placed in universities, small and medium-sized enterprises, and research institutes, the network aims at knowledge exchange across various disciplines, alongside professional and personal development of future leaders. Such an environment trains the researchers in communicating their results to different audiences, analysing large amounts of data and managing different projects. Collaboration is key and is one skill that is transferable to various work environments.

In the 2017 Career Tracking Survey of Doctorate Holders carried out by the European Science Foundation- Science Connect,[11] in partnership with nine universities and organisations, about 2000 respondents were asked to rate their competence at the end of their doctoral studies and the importance in their current job. The most important factors that came up were critical-analytical thinking, problem-solving, and effective communication.

"A lot of employers are becoming more aware of the fact that PhDs have a lot of experience in data gathering, analysis and management. It is in their interest to leverage this training", says Dr Verity Elston, head of career advice for PhDs and postdocs at Graduate Campus University of Lausanne, Switzerland. "With the automatisation of labour and the changing employment market,

the most important skills that a machine can't have, such as creativity, critical thinking, and adaptability are what PhD graduates can offer. I sense that as we move further into the twenty-first century, we will see an increasing demand for the kind of professional that a PhD is trained to be. Employers will see the value of having a natural talent to innovate, do things differently, and learn and adapt rapidly to changing environments," continues Dr Elston.

Applications: Putting your data and communication skills into action

Dr Rajaneesh Gopinath, who pursued an academic career in life sciences and bioinformatics data analysis, used his communication skills to transition first into a freelance scientific editor role, before taking up a business development role in the same biotech media company.

"My work is part writing and part market analysis as we report on biotech stocks, mergers, and acquisitions, and drug market analysis. So, my training as a scientist who churned data comes in handy for my current role," says Dr Gopinath.

Combining data and communication skills seems to be crucial for data scientists in general. When Kaggle surveyed about 7000 data scientists in 2017 and asked them about barriers at work,[12] the top four challenges were non-technical ones: lack of management/financial support, lack of clear questions to answer, results not used by decision-makers, and explaining data science to others. This again points towards the added value of highly skilled professionals who understand data and can communicate effectively.

"I used to do science communication on the side when I was doing my Ph.D./Postdoc and it was appreciated at all the job interviews. Communicating my thesis and writing other technical articles had already exposed me to various facets of science communication," says Dr Rebecca Alexander, data science professional at Voodoo.io. "In my current role, I write documentation or do presentations for different stakeholders at differing levels of expertise."

> "The best thing about data science and good communication skills is its transferability and relevance across industries."
>
> Dr Rebecca Alexander

Overall, it seems that the global job market is evolving rapidly, and while it is difficult to predict how it will shape up in the next decade, one thing is for sure: highly skilled professionals such as PhD graduates will add value to organisations due to their technical as well as interpersonal skills. In the life-science industry, digitalisation and data are crucial for achieving goals in precision medicine, personalised health and implementing technologies using blockchain in the future. Investing time in becoming data literate, along with improving on soft skills, can likely open many doors for any professional.

Conflicts of interest

The author declares no conflicts of interest.

References

1. Becoming A Data-Driven CEO | Domo [Internet]. [cited 2020 Feb 29]. Available from: https://www.domo.com/solution/data-never-sleeps-6
2. Baby K, Ravikumar A. Big data: An ultimate solution in health care. Int J Comput Appl. 2014;106:28–31.
3. Regulating the internet giants – The world's most valuable resource is no longer oil, but data | Leaders [Internet]. The Economist. [cited 2020 Feb 29]. Available from: https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data
4. Data as the new currency | Deloitte Insights [Internet]. [cited 2020 Feb 29]. Available from: https://www2.deloitte.com/us/en/insights/deloitte-review/issue-13/data-as-the-new-currency.html
5. Is data the new currency? [Internet]. World Economic Forum. [cited 2020 Feb 29]. Available from: https://www.weforum.org/agenda/2015/08/is-data-the-new-currency/
6. 3 top take-aways from the Gartner Chief Data Officer Survey [Internet]. 2018. [cited 2020 Feb 29]. Available from: https://www.gartner.com/smarterwithgartner/3-top-take-aways-from-the-gartner-chief-data-officer-survey/
7. Hybrid jobs: How new skills are rewriting the DNA of the job market [Internet]. Burning Glass Technologies. [cited 2020 Feb 29]. Available from: https://www.burning-glass.com/research-project/hybrid-jobs/
8. Gould J. How to build a better PhD. Nature. 2015 Dec;528(7580):22–5.
9. Education at a Glance 2019 | READ online [Internet]. OECD iLibrary. [cited 2020 Mar 1]. Available from: https://read.oecd-ilibrary.org/education/education-at-a-glance-2019_f8d7880d-en
10. Roach M, Sauermann H. The declining interest in an academic career. PLOS ONE. 2017; 12(9):e0184130.
11. ESF-Science Connect publishes outcomes of its career tracking survey of doctorate holders | News | CORDIS | European Commission [Internet]. [cited 2020 Mar 1]. Available from: https://cordis.europa.eu/article/id/135931-esfscience-connect-publishes-outcomes-of-its-career-tracking-survey-of-doctorate-holders
12. The State of ML and Data Science 2017 [Internet]. Kaggle. [cited 2020 Mar 1]. Available from: https://www.kaggle.com/surveys/2017
13. Berinato S. Data science and the art of persuasion [Internet]. 2019. [cited 2020 Mar 1]. Harvard Business Review. Available from: https://hbr.org/2019/01/data-science-and-the-art-of-persuasion

Author information

Menorca Chaturvedi, PhD, is a biomedical data analyst and a communications professional. She was the 2014 recipient of the EMWA Geoff Hall Memorial Scholarship and is currently based in Basel, Switzerland.

Digital identifiers in scientific publishing and e-health

Hamid Reza Khedmatgozar
Iranian Research Institute for Information Science and Technology (IRANDOC)
Tehran, Iran

Correspondence to:
Hamid Reza Khedmatgozar
Assistant Professor in Information Technology Management
Iranian Research Institute for Information Science and Technology (IRANDOC)
No. 1090, Enghelab Avenue
P.O. Box 13185-1371
Tehran, Iran
khedmatgozar@irandoc.ac.ir
h.khedmatgozar@gmail.com

Abstract
With the growth of information technology over the past decade, digital identifiers have been introduced for the unique and stable identification of digital objects in cyberspace. Digital identifiers have applications in many contexts, including scientific publications. In addition to describing their use for scientific publications, this article presents an additional potential application in the field of electronic health and invites experts and researchers to investigate it further.

I recently read an article published in this journal in 2013 entitled "AuthorAID: An international service and chance to serve" by Barbara Gastel.[1] Barbara states that too little attention has been paid to researchers in low- and middle-income countries, and that the scientific community needs to consider the results of their research. She mentions the growing gap in scientific publications between low-income countries and the rest of the world in terms of number and content of research articles, and the process for getting them published, as well as ongoing efforts to make scientific publications more accessible to readers everywhere in the world. Since the publication of her paper, developments have helped to reduce this gap and increase the availability of scientific publications. In this article, I will try to illustrate one of these developments and research areas relevant to it.

What is a digital identifier?

Today, many of our professional and leisure activities have moved to cyberspace. Many objects and entities in cyberspace (including persons, corporate bodies, animals, books, articles, journals, and songs) have a distinct identity and are called digital objects. How to deal with them is a big issue. In general, dealing with digital objects is based on two basic approaches: (1) archiving digital objects and related information (metadata) to organise and preserve them over time ("memorable web" approach), and (2) increasing the digital object's find-ability through a deep search on the web ("navigable web" approach). More recently, a third approach called the "identifiable web" or "digital identifier" approach has emerged.[2] This third approach is close to the memorable web approach with its focus on uniqueness and persistent identification of digital objects, but it also offers the main feature of the navigable web approach by focusing on a specific mechanism for instant online access to digital objects. In this article, I introduce research and applied aspects relating to this approach.

Currently, uniform resource locators (URLs) are used to identify and access digital objects. However, using URLs as identifiers carries risks, including lack of persistency (inability to access the object due to server migration or digital object transfer – "error 404") and violations of uniqueness (multiple URLs for a specific digital object). To overcome these risks, researchers have tried to create a unique identifier with long-term persistence. The main solution proposed is the use of digital identifiers instead of URLs.[3]

The digital identifier is a code assigned to a digital object in cyberspace that is completely unique, much like a fingerprint. There is a one-to-one relationship between the digital identifier and the identified digital object. Digital identifiers are issued by independent authorities, which enhances their reliability and accuracy. Each digital identifier acts as a link so that anyone clicking on it will either be redirected to a valid URL for the object or to metadata about the object, such as its creator, publisher, and format. This redirection is achieved by a mechanism called resolution (Figure 1).

The main functions and benefits of digital identifiers are unique and persistent identification on the web, permanent maintenance of location information (even if the digital object's location changes), standardisation and enrichment of metadata, facilitation of content searching, securing of copyright, increased traffic to digital objects, reduced costs, time savings for stakeholders, exploitation of new business and research areas, and increased revenues from digital objects. This digital identifier approach has important applications in scientific publications and research data. Other potential applications include interactive television, digital museums, virtual tourism, and e-learning.[3] Based on my own investigations, another potential application is electronic health (e-health).

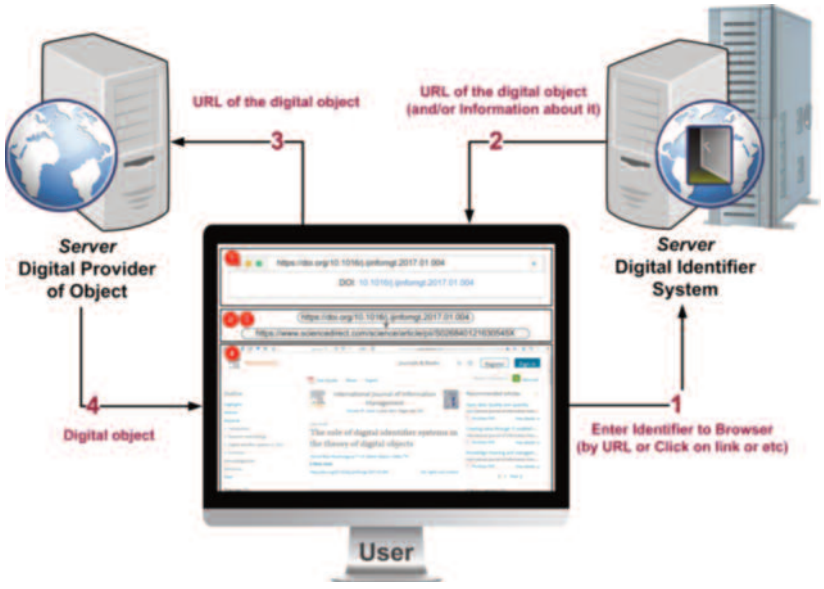


Figure 1. Example of a resolution mechanism for accessing a digital object through its digital identifier

A digital identifier is a piece of actionable code in cyberspace that works based on resolution. This mechanism has four steps. The first two steps involve redirecting the digital identifier to the location (URL) of the identified digital object and the last two steps involve providing access to this object via its URL. All of these steps are done in less than a second by your web browser.

Digital identifiers and scientific publications

In the context of scientific publications, digital identifiers and related systems can help stakeholders in the process of writing, editing, submitting, reviewing, and publishing papers in scientific journals. Some specific applications of digital identifiers are listed in Table 1.

Digital identifiers mean connecting to millions of other scientific articles via hyperlinks, with the digital identifier acting as the fingerprint of an article on the web. One of the most important applications of digital identifiers in scientific publications is the Digital Object Identifier (DOI) system and related registration agencies such as Crossref and DataCite.

DOI is the most well-known digital identification system. It is managed and controlled by the International DOI Foundation (IDF), a non-profit organisation. A DOI is permanently allocated to an object to provide a stable hyperlink that redirects to current information about that object, including its location. While information about an object can change over time, its DOI will not change. DOIs are allocated on behalf of the IDF by its registration agencies. The IDF currently has 10 registration agencies. As shown in Figure 2, DOI syntax consists of an indefinite character string including a prefix and a suffix. The prefix is a directory code that identifies the IDF as the directory and that is always 10 in DOI, followed by a registrant code, which is the unique code assigned by IDF registration agencies to the owner or publisher of the object being identified. The directory code and the registration code are separated by a full stop. After the prefix is a forward slash and then the suffix. This suffix is a unique code containing alphanumeric strings assigned by the registrant to identify the object. This suffix must be unique to the registrant and can include other identifiers such as the ISBN, ISSN, or serial number. There are no operational restrictions on the length of a DOI or any of its components, and identifiers up

Table 1. Specific applications of digital identifiers

Application	Explanation
Enabling unique and persistent identification	Digital identifiers enable the creation of unique, persistent, automated links between digital objects such as journals, publishers, editors, authors, and even components of an article such as a table or a dataset and remove human error in maintaining the links between them. They also enable automatic and reliable retrieval of information about articles.
Facilitating acceptance by indexers	In their initial evaluations of journals, many indexers, such as International Scientific Indexing, PubMed, Scopus, and Directory of Open Access Journals, ask whether the journal uses digital identifiers. Although not stated as a requirement for indexing, journals that use digital identifiers generally find it much easier to get accepted by indexers. Moreover, it has been claimed that if a journal is indexed by one of the major indexers and starts using digital identifiers, then its impact factor can increase greatly within a year.[9]
Automating the citation of articles, journals, and authors	Digital identifiers can be used to automatically monitor the level of influence of an article, author, or journal.
Increasing visibility	A big challenge for every journal and author is to bring the content of their article to the largest possible number of readers. Information about articles with digital identifiers is recorded in multiple reputable databases, which increases visibility of the articles in search engines. This increase in visibility is beneficial to journals and authors and also helps readers to find journal articles.
Preventing plagiarism	Using digital identifiers helps to maintain copyright by providing direct access to the official links for articles. iThenticate, the most reputable plagiarism detection service, only includes in its database published articles with digital identifiers. If a journal article has a digital identifier, it will automatically be entered into the database. From the time a journal subscribes to iThenticate, iThenticate will notify the journal and its publisher if another article with even the slightest resemblance to one of their articles is published.

to 4 GB can be assigned. To use the DOI system and access the DOI directory, https://doi.org/ must be placed before the DOI syntax.[4]

The mission of Crossref, an independent nongovernmental company launched in 2000 and now run by research publishers, is to facilitate access, citation, and reuse of articles and books as research outputs. Through the DOI, Crossref enables researchers to easily switch from article to article at the click of a mouse, even if the articles are in different journals. The company also has other services based on the DOI, such as Similarity Check, Crossmark, and Crossref Metadata Search. Similarity Check is a service developed by iThenticate[5] that allows publishers to check the authenticity and originality of the articles that are submitted to them. Crossmark can be used to check whether or not you are using an up-to-date version of an electronic article. And Crossref Metadata Search automatically provides the information needed by some indexers without journal interference. Figure 3 shows an example of easily switching from article to article. Publishers of scientific content, especially books and journals, can access Crossref to use its services, including DOIs.[6]

> Optimal use of the specific features of digital identifiers, including persistency and uniqueness, can help to reduce the gaps between rich and poor in information and health services in the public health field.

Researchers have long wanted a better system for sharing data, as well as data archiving tools for reusing data in future studies. The need to create and manage a process that provides continuous and reliable access to data led to the establishment of DataCite in 2009 with the aim of assigning a unique identifier to scientific datasets. To this end, DataCite collaborates with a number of research libraries worldwide. According to DataCite, data should be cited in the same way that articles and books are cited. Citation allows for verification and reuse of data and allows researchers to be identified and rewarded for their contributions based on the impact of their dataset. Researchers and organisations can contact a DataCite member to obtain a DOI for each of their datasets. For example, researchers at the ResearchGate[7] scientific social network can share their datasets, receiving DOIs through DataCite.[8]

However, in developing countries, scientific publishers use DOIs significantly less than in developed countries, often due to lack of awareness and low financial capacity. As a result, the scientific outputs of researchers in developing countries are read and cited less than those in developed countries. Also, the limited citations these scientific outputs receive cannot be tracked and displayed to researchers because DOIs have not been used. This is the most important gap that the wider use of DOIs could help reduce and is consistent with the research results of Boudry and Chartron.[10] They systematically studied the use of the DOI in articles indexed in PubMed and concluded that the DOI is more frequently used by journals in developed countries than in developing countries and in international journals than in regional journals. As well as demonstrating the use of the DOI in increasing global knowledge sharing, they propose that, in order to further increase and improve sharing,

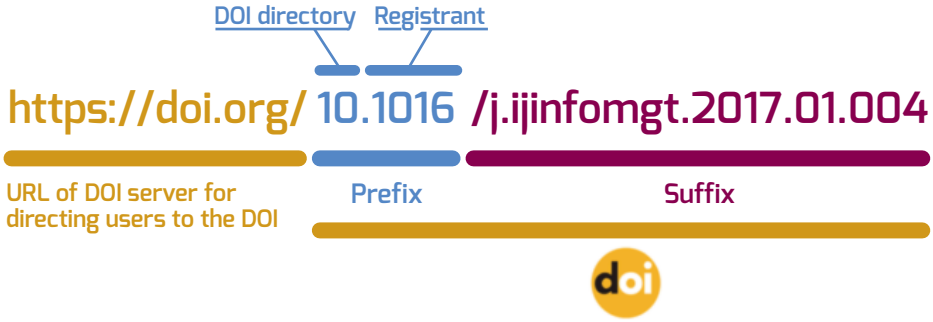


Figure 2. Syntax of a DOI
DOI syntax has two parts: prefix and suffix. The prefix consists of two parts, the directory and registrant, and is assigned by a DOI registration agency. The suffix is assigned by the owner or publisher of the identified object. In order to use this syntax to implement resolution, the web address (https://doi.org/) must be placed before the prefix.

Figure 3. Example of easily switching from article to article using DOIs
When you read an article, you see that another article has been referenced. You can easily access that reference by clicking on its DOI. You can also click on the CrossMark icon to check whether you are viewing the latest version of the article.

necessary measures must be undertaken to allocate this identifier to journals in developing counties.

Digital identifiers and e-health

In today's world, health, technology, and business have found a strong link in the concept of e-health. The dominant term in the technology aspect of this concept is the internet. [11] Systems such as electronic health records and health management information systems (HMISs) also fall under this concept. An electronic health record is a longitudinal electronic record of a patient's health information generated as a result of one or more encounters in a care delivery setting.[12] An HMIS is a system whereby health data are recorded, stored, retrieved, and processed to improve decision-making.[13]

At the macro level, the literature in this area indicates that in order to achieve universal health coverage, a unique health identifier is essential to identify individuals in a country, especially patients.[14] Efforts have been made in different countries to assign identifiers to patients. For example, Mills and colleagues[14] have compared efforts in the United Kingdom, Slovenia, South Korea, and Thailand. They point out that, in some countries, there are problems with the lack of access to identifiers at all times, as well as the existence of local identifiers and the lack of links between local and national identifiers. Studies in other countries such as Cambodia, Laos, and Myanmar,[15] as well as Nigeria,[16] point to other similar problems. It seems that the benefits of a digital identifier solution, including guaranteed stability and metadata, can be used to address some of these problems. A joint study of unique health identifiers and digital identifiers could be of interest to researchers.

One of the basic processes in the field of health in general and e-health in particular is the unique identification of all kinds of non-patient stakeholders and objects, such as nurses, hospitals, medicines, therapists, and pharmacies. If these stakeholders and objects are uniquely identified, it is possible to link them and their data, connect different HMISs, and thereby monitor the stakeholders and objects. For example, Sensmeier and colleagues[17]

emphasise the need for the unique digital identification of nurses in both the health/treatment process and digital information systems in order to monitor their actions and evaluate their performance. Given the advantages of digital identifiers, it seems they would be suitable for uniquely and persistently identifying and monitoring non-patient stakeholders and objects.

Health research continues to generate a large number of valuable datasets. An example is the field of genetics and in particular genomics (most recently in relation to COVID-19). In order to have reliable and permanent global access to these datasets, more attention needs to be paid to assigning digital identifiers to them. The work at DataCite can be a precursor to these efforts, which should be of great interest to health professionals and researchers.

Conclusion

In the field of scientific publications, digital identifiers simultaneously provide unique, persistent identification and persistent access to scientific outputs such as articles and datasets. Through the use of digital identifiers in low- and middle-income countries, research outputs will become more accessible to the world at large, thereby reducing the research gap between richer and poorer nations. So, if you are a reader, author, reviewer, or editor of scientific research, I suggest you become more familiar with digital identifiers and related services.

The use of digital identifiers in e-health is a new research subject of potential interest to researchers in both information technology and health. It can also be considered as a business opportunity for practitioners in these areas. In my opinion, optimal use of the specific features of digital identifiers, including persistency and uniqueness, can also help to reduce the gaps between rich and poor in information and health services in the public health field. For example, digital identification of a person living in a slum area or low-income country combined with persistent access to his or her health profile can enable centralised monitoring of all health services provided to the person by all stakeholders. Decision-makers can then exploit the persistent and unique identification of digital objects including patients, stakeholders, and datasets to identify health service gaps between rich and poor areas and work to reduce them.

Conflicts of interest

The author declares no conflicts of interest.

References

1. Gastel B. AuthorAID: An international service and chance to serve. Med Writ. 2013;22(4): 284–7. doi: 10.1179/2047480613Z.000000000151
2. Khedmatgozar HR, Alipour-Hafezi M. The role of digital identifier systems in the theory of digital objects. Int J Inform Manage. 2017;37(3):162–5. doi: 10.1016/j.ijinfomgt.2017.01.004
3. Khedmatgozar HR, Alipour-Hafezi M. A basic comparative framework for evaluation of digital identifier systems. J Digit Inf Manag. 2015;13(3):190-7.
4. International DOI Foundation (IDF). DOI Handbook. 2020 [cited 2020 Apr 9]. Available from: https://www.doi.org/10.1000/182
5. iThenticate: Plagiarism Detection Software. 2020 [cited 2020 Apr 9]. Available from: http://www.ithenticate.com/
6. Crossref. 2020 [cited 2020 Apr 9]. Available from: https://www.crossref.org/
7. ResearchGate. 2020 [cited 2020 Apr 9]. Available from: https://www.researchgate.net/
8. DataCite. 2020 [cited 2020 Apr 9]. Available from: https://datacite.org/
9. Braile DM. After the impact factor, the DOI. Rev Bras Cir Cardiovasc. 2011; 26(3):I-II. doi: 10.5935/1678-9741.20110001
10. Boudry C, Chartron G. Availability of digital object identifiers in publications archived by PubMed. Scientometrics. 2017;110(3):1453–69. doi: 10.1007/s11192-016-2225-6
11. Oh H, Rizo C, Enkin M, Jadad A. What is eHealth (3): A systematic review of published definitions. J Med Internet Res. 2005;7(1):e1. doi: 10.2196/jmir.7.1.e1
12. Health Information and Management Systems Society. EHR: Electronic health record. 2020 [cited 2010 Mar 20]. Available from: https://www.himss.org/electronic-health-records
13. Endriyas M, Alano A, Mekonnen E, et al. Understanding performance data: Health management information system data accuracy in Southern Nations Nationalities and People's Region, Ethiopia. BMC Health Serv Res. 2019;19(1):175. doi: 10.1186/s12913-019-3991-7
14. Mills S, Lee JK, Rassekh BM, et al. Unique health identifiers for universal health coverage. Journal of Health, Population and Nutrition. 2019;38 (1):22. doi: 10.1186/s41043-019-0180-6
15. Stahl M, Roth S, Thorell L, Parry J. On the road to universal health coverage: every person matters. ADB Briefs. 2016; No. 56:1–8.
16. Chukwu E. The case for a unique digital patient ID scheme in Nigeria. J Health Med Informat. 2017;8(3):267–71. doi: 10.4172/2157-7420.1000267
17. Sensmeier J, Androwich I, Baernholdt M, et al. Demonstrating the value of nursing care through use of a unique nurse identifier. Online J Nurs Inform. 2019;23(2).

Author information

Hamid Reza Khedmatgozar is an Assistant Professor of Information Technology Management at the Iranian Research Institute for Information Science and Technology (IRANDOC) in Tehran, Iran. His research interests revolve around IT adoption, digital identifier systems for information objects, e-business models, risk management, neural networks, and multiple-criteria decision-making.

Winners of the Geoff Hall Scholarship Essay Competition

Dear all,

You will be reading this during unprecedented quarantine or lockdown (depending on your geographical location). My thoughts and prayers are with you all.

The Geoff Hall Scholarships are given in honour of a former president of EMWA. Geoff was a very special person, an extremely valued member of EMWA, and a very good friend to many EMWA members. He firmly believed that the future of EMWA lies in our new and potential members, and so it's a very fitting legacy that we have the Scholarship Awards in his memory. The scholarships are awarded annually on the basis of an essay competition, and the title of this year's essay was "How would you go about identifying a predatory journal?" This year's scholarship winners were Adriana Rocha and Petal Smart.

FOR CORRESPONDENCE

Lisa Chamberlain James
Lisa@trilogywriting.com

Adriana Rocha has a degree in biochemistry from Portugal, which was followed by a PhD in medical neurosciences in Germany. After a postdoc in the USA, she decided to leave academic research and transition into industry. She is now a freelance medical writer.

Petal Smart is a veterinary surgeon by training. Over the past 5 years, she has been a medical/science editor serving primarily non-native English-speaking authors. She has a keen interest in regulatory affairs as they relate to medical devices, both those intended for human use and those intended for veterinary use.

Adriana's and Petal's winning essays are presented below, and we wish them the very best at the start of their very promising medical writing careers. For those of you inspired to pick up your laptop, and are looking for something to fill your time during quarantine, the essay title for this year's upcoming competition is "Do you have what it takes to be a medical writer? Discuss three attributes or skills that best qualify one to be a medical writer".

I hope to read your essays soon and stay safe all until we see each other at the next EMWA conference.

Bestest,
Lisa

Investigate: What is a predatory journal?

Why it is a problem and what are the dangers?

For the last 15 years, predatory journal numbers have exploded.[1] Such journals are not legitimate scientific publications and their business model consists of for-profit publishing, where the peer-review process is mostly non-existent and there is no concern for the scientific accuracy of the published content. They are an obvious danger for the integrity of the scientific method and conflating predatory journal articles with actual scientific content is not only a waste of knowledge, time and resources: it undermines the value of legitimate published scientific research.[1–3]

But if they are so unethical, why have these journals been booming? Unfortunately, most universities and institutions evaluate researchers by their publication output, forcing them to publish often and regularly to maintain their professional standing. In this context, a journal that promises a quick publication process can be very attractive, especially to less-experienced researchers who are not aware of the risks.[1-3]

Sometimes even experienced researchers can unknowingly submit an article to a predatory journal. When the authors realise they are dealing with a predatory publication and wish to retract the article, the journals will usually charge a retraction fee.[4] A few publications even list respected professionals as board members who have never agreed to do so and only discovered it when contacted by a fellow researcher.[5]

On the other end of the spectrum, some researchers will deliberately use predatory journals and, for a fee, increase their number of publications solely to advance their career with no concern for ethics and scientific accuracy.[3]

In their quest to denounce predatory journals, some researchers have shown how easy it is to scam your way to a publication. In 2013, John Bohannon had his fake (and flawed) research article accepted for publication by 157 of 304 open access journals, contingent upon payment of publication fees.[6] Katarzyna Pisanski and colleagues also showed how to easily join an editorial board of a predatory journal. They created a profile of a fictitious scientist named Anna O. Szust and applied on her behalf to the editorial boards of 360 journals. Oszust is the Polish word for fraud and the scientist's CV was clearly unqualified for an editorial role. In many cases, they received a positive response within days and even hours of application. Four titles immediately appointed Szust editor-in-chief (!) and at least a dozen journals appointed Szust as an editor, conditional upon some form of payment.[7]

All these examples illustrate the lack of ethical practices and the for-profit nature of predatory journals. But new journals are being

launched every week, how can we know if they are legitimate or not? How do we safely recognise a predatory journal?

Identify: How to recognise a predatory journal

In the face of such an extensive number of fraudulent publications, many organisations are recognising the problem and advising their members on how to better recognise and identify a predatory journal.[1,2,8,9]

Usually, a predatory journal will send emails to researchers requesting article submission and promising a fast and easy publication process. Afterwards, the fees charged are suspicious (too high or too low) and very often payable upon article submission. Further investigation reveals that they often have a legitimate-sounding name (in fact just a variation of an original legitimate publication), and possibly an unprofessional website, with fake addresses and grammatical errors. Upon further inspection, it is revealed the journal is not indexed on recognised citation systems such as PubMed[10] nor listed within an accredited online directory such as the Directory of Open Access Journals (DOAJ),[11] The DOAJ lists legitimate open access journals: it grants some journals the DOAJ seal, a mark of certification for open access journals for the achievement of a high level of openness, adhering to best practices, and having high publishing standards. As of September 2019, it listed 13,776 journals in 130 countries. In any case, it should be mentioned that the DOAJ is not a comprehensive list of all legitimate open access journals and a journal that is not listed should not be assumed to be predatory. In addition, the listing itself does not guarantee legitimacy, but the DOAJ has a routine mechanism for users to notify DOAJ if they find a journal with questionable practices.[11]

One very useful tool that indexed fraudulent publications was Beall's list of predatory publishers. From 2011 to January 2017, Jeffrey Beall, a librarian and associate professor at the University of Colorado, compiled annual lists of potential, possible, or probably predatory open access journals.[12] As of January 3, 2017, it listed 1155 predatory publishers and 1294 predatory journals.[2] However, on January 17, 2017, Beall's website was shut down for unclear reasons.[13]

Inform: Where to find more up-to-date information Share it with other professionals

As mentioned previously, new journals are being constantly created and ultimately it is the researcher's responsibility to ascertain which are legitimate or predatory. Not one method is full-proof: the best process involves a combination of techniques, such as applying the criteria for predatory journals, seeing if they are listed on the DOAJ or other online directories and contact other senior colleagues to see if they have heard of the journal.

A useful tool to systematise this analysis comes from "Think. Check. Submit" – an online checklist developed by a coalition of scholarly publishing organisations.

A useful tool to systematise this analysis comes from "Think. Check. Submit." – an online checklist developed by a coalition of scholarly publishing organisations. Again, this is merely a way to better guide your research into a particular journal and incorporate all the criteria mentioned previously.[14]

Ultimately, *you* have an individual, scientific and ethical responsibility to identify and avoid predatory journals and only publish in legitimate publications. In this era of fake news, it is up to us – researchers, medical writers, editors, and respective organisations – to inform and educate, so that science is peer-reviewed, reliable, and rigorous.[14]

References

1. American Medical Writers Association, European Medical Writers Association, International Society for Medical Publication Professionals. AMWA-EMWA-ISMPP joint position statement on predatory publishing. Curr Med Res Opin. 2019 Sep;35(9):1657–8.
2. Laine C, Winker MA. Identifying predatory or pseudo-journals. World Association of Medical Editors (2017). Available from: http://www.wame.org/identifying-predatory-or-pseudo-journals
3. Moher D, Shamseer L, Cobey KD, et al. Stop this waste of people, animals and money. Nature. 2017 Sep;549(7670):23–5.
4. Cobey K. Illegitimate journals scam even senior scientists. Nature. 2017 Sep;549(7670):7.
5. Cobey KD, Lalu MM, Skidmore B, Ahmadzai N, Grudniewicz A, Moher D. What is a predatory journal? A scoping review. F1000 Res. 2018 Jul;7:1001.
6. Bohannon J. Who's afraid of peer review? Science. 2013 Oct;342(6154):60–5.
7. Sorokowski P, Kulczycki E, Sorokowska A, Pisanski K. Predatory journals recruit fake editor. Nature. 2017 Mar;543(7646): 481–3.
8. International Committee of Medical Journal Editors. Responsibilities in the submission and peer-review process. Available from: http://www.icmje.org/recommendations/browse/roles-and-responsibilities/responsibilities-in-the-submission-and-peer-peview-process.html
9. Cabell's International Blacklist criteria. Available from: http://www2.cabells.com/blacklist-criteria
10. PubMed, Available from: https://www.ncbi.nlm.nih.gov/pubmed
11. Directory of Open Access Journals (DOAJ). Available from: https://doaj.org/
12. Beall J. Beall's list of predatory publishers (2016) Scholarly Open Access. Available from:https://web.archive.org/web/20170113114519/https://scholarlyoa.com/2016/01/ 05/bealls-list-of-predatory-publishers-2016/
13. Chawla DM. (2017) Mystery as controversial list of predatory publishers disappears. Science. DOI:10.1126/science.aal0625
14. Think. Check. Submit. Available from: https://thinkchecksubmit.org/

Adriana Rocha
adriana.duarte.rocha@gmail.com

How would you go about identifying a predatory journal?

You are anxious to make progress – in your career, your research – the faster, the better. However, the publish-or-perish "jungle" is dark and murky. You can barely see what lies ahead. Your instincts remind you that "they" are out there, to get you – by any means necessary. You are the prey and "they" are the predatory journals. They want your blood, sweat, and hard-earned reputation for less-than-noble purposes.

Cohen et al.[1] define a predatory journal as "an exploitative open-access academic publishing business model that involves charging publication fees to authors without providing the editorial and publishing services typically associated with legitimate journals". Although early-career researchers are more likely to be preyed upon, more experienced scientists are not exempt, and may be targeted to serve as editors or reviewers.

Scientists should think of predatory publishers as similar to counterfeit money. They may appear to be authentic, but you need to look very closely with an analytical eye to determine their legitimacy.

Because predatory journals may go to extreme lengths to convince you of their supposed legitimacy, you need to be alert to detect the subtle signs of deception. If something appears to be questionable, it probably is. Thus, researchers should also trust their instincts. Although the following guidelines are by no means exhaustive or even fool-proof, you should consider the following:

Website

Is there a legitimate, up-to-date website for the journal? Are the journal name and website URL unique, or do they very closely resemble those of another well-established publication? Are there spelling errors on the website of the journal? Does the website resemble a sales pitch to authors, or does it appeal to its target audience?

Contact details

Does the website clearly list the editorial staff and their full contact details (email, telephone, physical mailing address)? Does the last part of the email address correspond with the official website of the journal, or is it a freely available email address from one of the popular email providers (e.g., Gmail, Yahoo, or Hotmail)? For instance, if their website is legitimatepublisher.com, do you have to email the editor at editor@legitimatepublisher.com, or editor.predatorypublisher@gmail.com?

You should try to contact the editor with a pertinent question and see whether you receive a response. In some cases, there may be glaring disparities. For instance, is the editorial address for the *European Journal of XYZ* really based anywhere in Europe?

Editorial staff

Who are the editor-in-chief and managing editor? What are their backgrounds? Can you verify their

credentials? Do their credentials match the subject matter of the journal?

Even if you are investigating one particular journal, do a quick check of the editorial staff of other journals published by the same publisher. Is the same editor responsible for various journals in unrelated subjects? Unlike language editors and copy editors (who are quite capable of checking manuscripts in various subjects for errors in syntax, grammar, spelling, accuracy, etc.), journal editors are usually specialists in their particular field. Therefore, their background should match the subject matter of the journal for which they are listed as an editor.

Does the editor have a list of publications themselves? The purpose of this evaluation should not be to discriminate against any particular scientist. However, it is more likely that an experienced scientist would be qualified to assume an editorial or leadership role at an academic publication.

Can the publications and credentials of the editors be searched for and found online? In some cases, the editors listed may be fictitious names. In other cases, legitimate scientists may not even be aware that they are listed as editors on the websites of some predatory journals.

Peer review

Is this process reasonably rigorous? A legitimate journal would want to impose a certain measure of scientific rigor to ensure quality control. Is the length of time or the process of peer review remarkably short? Is the process of submission for review incredibly easy or "too good to be true?"

How many reviewers are typically selected? One, two, three, or more? Are any reviewers selected at all? Is the process transparent and is this information readily available in the information for authors or submission guidelines? Are you able to recommend any reviewers? Are there processes in place to respond to reviewers' comments and suggestions and resubmit the manuscript?

Although predatory journals may claim to conduct peer review and mimic the structure of legitimate journals, they publish all or most submitted material without external peer review. They often disregard policies "advocated by organisations such as the World Association of Medical Editors (WAME), the Committee on Publication Ethics (COPE), the International Committee of Medical Journal Editors (ICMJE), and the Council of Science Editors (CSE) regarding issues such as archiving of journal content, management of potential conflicts of interest, handling of errata, and transparency of journal processes and policies including fees."[2]

> Because predatory journals may go to extreme lengths to convince you of their supposed legitimacy, you need to be alert to detect the subtle signs of deception. If something appears to be questionable, it probably is.

Impact factor

Are there processes in place to monitor the number of times articles have been cited? (Keep in mind, it may take at least a year or two for a new journal to establish this.) Does the journal itself cite a reputable impact metric (e.g., Web of Science, CrossRef, or Altmetric)? Does the website clearly state how the articles are processed and archived for future citation or referencing?

Copyright

Who retains rights to the published articles, and is this clearly stated in the information for authors or submission guidelines? In an open access model, the authors usually retain rights, as opposed to the publisher.

Other general considerations

Although scientists have referred to Beall's list for a number of years, because it was originally compiled and maintained by one individual, its reliability has been called into question, and it now exists only as an online archive.[2]

Other resources, such as the "Think. Check. Submit" initiative,[3] and the Directory of Open Access Journals are useful resources that should be consulted when preparing to submit research for publication.

Keep in mind that predatory journals are money-making enterprises. Therefore, when confronted with predators in the "jungle", think of their driving force, trust your "survival" instincts, and be alert to their deceptive practices.

References

1. Cohen AJ, Patino G, Kamal P, et al. Perspectives from authors and editors in the biomedical disciplines on predatory journals: Survey study. J Med Internet Res. 2019;21(8):e13769.
2. Laine C, Winker MA. Identifying predatory or pseudo-journals. World Association of Medical Editors. February 15, 2017. [cited 2019 Sep 30]. Available at: https://www.wame.org/identifying-predatory-or-pseudo-journals
3. Think. Check. Submit. Choosing the right journal for your research. [cited 2019 Sep 30]. https://thinkchecksubmit.org/

Petal Smart
petalsmart@pm.me

News from the EMA

SECTION EDITOR

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

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

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

Ten recommendations to unlock the potential of big data for public health in the EU

January 20, 2020 – The joint Big Data Task Force of European Medicines Agency (EMA) and the Heads of Medicines Agencies (HMA) proposes ten priority actions for the European medicines regulatory network to evolve its approach to data use and evidence generation, in order to make best use of big data to support innovation and public health, in a report published today.

Big data are extremely large, rapidly accumulating datasets captured across multiple settings and devices, for example through wearable devices, electronic health records, clinical trials, or spontaneous adverse reaction reports. Coupled to rapidly developing technology, big data can complement the evidence from clinical trials and fill knowledge gaps on a medicine, and help to better characterise diseases, treatments and the performance of medicines in individual healthcare systems. The rapidly changing data landscape forces regulators to evolve and change the way they access, manage, and analyse data and to keep pace with the rapid advances in science and technology.

The report makes several recommendations out of which ten are viewed as priorities. The most ambitious of these top ten recommendations is the establishment of an European Union (EU) platform to access and analyse healthcare data from across the EU (Data Analysis and Real World Interrogation Network, or DARWIN). This platform would create a European network of databases of verified quality and content with the highest levels of data security. It would be used to inform regulatory decision-making with robust evidence from healthcare practice.

Other recommendations are intended to enhance guidance and resources within the EU regulatory network for data quality and data discoverability (choice of key metadata) and to build up computing and analytical capacity. The joint task force advises to develop the skills to process and analyse big data within the network through training to enhance the capacity of regulators to assess applications for the authorisation of medicines that use big data sources as part of the evidence on benefits and risks. It proposes to establish a learning initiative to track and review outcomes of these types of submissions.

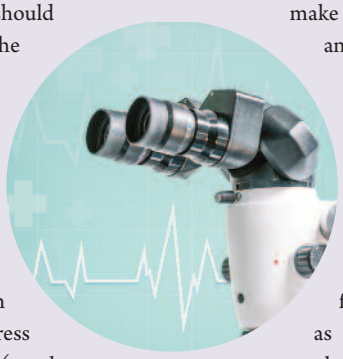
The report also emphasises the need to ensure data are managed and analysed within a secure and ethical governance framework, and in active dialogue with key EU stakeholders including patients, healthcare professionals, industry, health technology assessment bodies (HTAs), payers, device regulators, and technology companies. All these activities should be done in collaboration with international initiatives on big data.

Established in 2017, the HMA – EMA Joint Big Data task force is composed of experienced medicines regulators and data experts appointed by national competent authorities, EMA, and the European Commission. The first phase of its work – published in early 2019 – reviewed the landscape of big data and identified opportunities for improvements in the operation of medicines regulation. Published today, the practical suggestions made in the second phase of its work aim to inform strategic decision-making and planning by the HMA and EMA and to contribute to the upcoming EU Network Strategy to 2025.

Key principles for the use of electronic product information (ePI) for EU medicines

January 29, 2020 – EMA, the HMA of EU Member States, and the European Commission (EC) have published today key principles outlining a harmonised approach to develop and use electronic product information (ePI) for human medicines across the EU.

The product information of a medicine includes the package leaflet for patients and the summary of product characteristics (SmPC) for healthcare professionals. These documents accompany every single medicine authorised in the EU and explain how it should be prescribed and used. The package leaflet is provided in the medicine's box and can also be found, often as a pdf document, on the websites of EU regulators. However, digital platforms open additional possibilities to disseminate the product information electronically. This can address some of the current limitations (e.g. the current PI is not interoperable with other electronic health systems such as e-prescription and electronic health records) and better meet patients' and healthcare professionals' needs for accessible, trustworthy, and up-to-date information on medicines available at the right time.



The ePI initiative was launched to support the digital transformation of healthcare across the EU, and the commitment laid out by the EC to prioritise innovations that will empower citizens and build a healthier society. It is also in line with EMA's current digitalisation efforts aiming to make best use of available resources and prepare for future challenges.

The key principles describe the benefits ePI can deliver for public health and the efficiencies it may introduce in regulatory procedures. They explain how ePI will comply with the existing legislative framework: it will be provided as open access information that complements the paper package leaflet. They also outline a flexible, harmonised approach to implementation across the EU, and describe how ePI will work in the EU's multilingual environment and will interact with other ongoing digital initiatives at EU and global level.

The key principles derive from extensive discussions and consultations carried out in 2018 and 2019 by EMA, HMA and the EC with representatives of all stakeholder groups concerned, from patients, healthcare professionals, and regulators to the pharmaceutical industry. In particular, during a public consultation that took place from January to July 2019, 71 contributions from all stakeholder groups were received, including over 500 comments which were considered for the final version. A summary of the main points raised in the consultation and the submissions were also published today.

The key principles were endorsed at the end of 2019 by EMA's Management Board and by the HMA. They are now expected to be followed by all parties involved in the process of developing and implementing ePI for medicines across the EU.

First oral GLP-1 treatment for type 2 diabetes

January 31, 2020 – The EMA's human medicines committee (CHMP) has recommended granting a marketing authorisation in the EU for Rybelsus (semaglutide) for the treatment of adults with insufficiently controlled type 2 diabetes to improve glycaemic control as an adjunct to diet and exercise. It is the first glucagon-like peptide (GLP-1) receptor agonist treatment – a class of non-insulin medicines for people with type 2 diabetes – developed for oral use, providing patients with another option to treat the disease without injections.

Type 2 diabetes is a disease in which the pancreas does not make enough insulin to control the level of glucose in the blood or when the body is unable to use insulin effectively. Most people with diabetes have this form of diabetes. Possible complications of diabetes include heart attack, stroke, kidney failure, leg amputation, vision loss, and nerve damage.

The active substance in Rybelsus, semaglutide, acts in the same way as the incretin hormone GLP1: it reduces blood glucose by stimulating pancreatic secretion of insulin and lowering the secretion of glucagon (a hormone that works to raise blood sugar concentration) when blood sugar is high.

The safety and efficacy of Rybelsus were studied in eight clinical trials that included patients at various stages of the disease. In three of these studies, Rybelsus was compared to a placebo. In the development programme, it was either used on its own, added to the standard treatment or compared to an injection treatment of its same class (GLP-1 receptor agonist).

The most common side effects observed during the clinical trials were gastrointestinal side effects, such as nausea and diarrhoea. Hypoglycaemia may occur when used in combination with insulin or sulphonylurea.

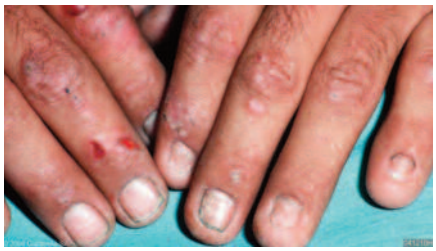
The opinion adopted by the CHMP is an intermediary step on Rybelsus's path to patient access. The CHMP opinion will now be sent to the EC for the adoption of a decision on an EU-wide marketing authorisation. Once a marketing authorisation has been granted, decisions about price and reimbursement will take place at the level of each Member State, taking into account the potential role/use of this medicine in the context of the national health system of that country.

First treatment for acute hepatic porphyria: Use of small interfering RNA

January 31, 2020 – EMA's CHMP has recommended granting a marketing authorisation in the EU for Givlaari (givosiran), the first treatment for acute hepatic porphyria (AHP) in adults and adolescents aged 12 years and older.

Acute hepatic porphyria is a rare genetic condition in which patients lack certain enzymes needed to produce haem, a basic structure of haemoglobin that binds to oxygen and is characterised by an accumulation of porphyrins in the body to toxic amounts. This can cause attacks of severe abdominal pain, vomiting, and nervous system disorders, such as seizures, depression, and anxiety. AHP is life-threatening due to the possibility of paralysis and respiratory arrest during attacks.

The new active substance givosiran is made of a short, synthetic strand of genetic material called 'small interfering RNA' that has been designed to interfere with the production of an enzyme involved in an early step in making haem. By blocking this early step of haem production in patients with AHP, the medicine is expected to prevent the next steps which produce substances that accumulate in the body and cause the



symptoms of the disease.

There are no approved treatments that directly ameliorate or prevent chronic symptoms experienced by many AHP patients and no approved treatments to reduce the risk of attacks. Intravenous hemin, a human blood-derived haem formulation, is the only therapy currently approved for the treatment of acute attacks. However, it is not approved as a chronic treatment to prevent attacks. Additional treatments include painkillers and antiemetics (to treat nausea and vomiting), chemically-induced menopause with hormonal suppression therapy, and liver transplantation.

The benefits and safety of Givlaari were demonstrated in a phase III clinical study which enrolled 94 patients with AHP who experienced at least two attacks in the past six months. Data from the study showed that the treatment resulted in a significant decrease of annual attacks, less pain, and an improved quality of life.

At the time of designation, AHP affected approximately 0.1 in 10,000 people in the EU and Norway, Iceland, and Liechtenstein, which makes it a rare disease. This product was designated as an orphan medicine during its development. At the time of approval, orphan designations are reviewed by EMA's Committee for Orphan Medicinal Products (COMP) to determine whether the information available to date allows maintaining the medicine's orphan status and granting the medicine ten years of market exclusivity.

Since Givlaari addresses an unmet medical need, it benefited from PRIME, EMA's platform for early and enhanced dialogue with developers of promising new medicines. This interaction led to a more robust application package to demonstrate the medicine's benefits and risks, which allowed the accelerated assessment of Givlaari in 150 days.

Guidance to sponsors on how to manage clinical trials during the COVID-19 pandemic

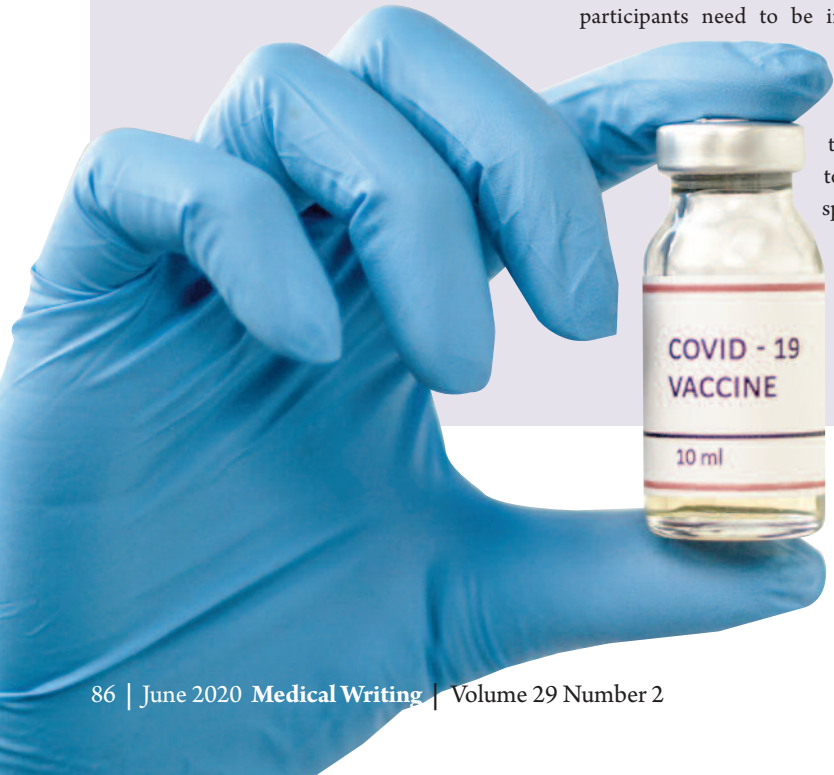
March 20, 2020 – The EC, the EMA and national HMAs have published new recommendations for sponsors on how to manage the conduct of clinical trials in the context of the coronavirus disease (COVID-19) pandemic. The impact of the pandemic on European health systems and more broadly on society, will make it necessary for sponsors to adjust how they manage clinical trials and the people who participate in these trials.

The guidance provides concrete information on changes and protocol deviations which may be needed in the conduct of clinical trials to deal with extraordinary situations, e.g. if trial participants need to be in self-isolation or quarantine, access to public places (including hospitals) is limited due to the risk of spreading infections, and healthcare professionals are being reallocated.

This guidance includes a harmonised set of recommendations, to ensure the utmost safety of trial participants across the EU while preserving the quality of the data generated by the trials. It also advises how these changes should be communicated to authorities.

There is specific advice on the initiation of new clinical trials for treatments of COVID-19, and in particular on the need for large, multinational trial protocols. This is in line with the call issued by EMA's CHMP for robust trial methodology in clinical trials for potential COVID-19 treatments or vaccines.

In the EU, clinical trials are authorised and supervised at national level. Sponsors are advised to also check whether there might be specific national legislation and guidance in place to complement or in some cases to take priority over this new guidance.



COVID-19: How EMA fast-tracks development support and approval of medicines and vaccines

May 4, 2020 – As researchers race to develop vaccines and therapeutics against COVID-19, EMA has published an overview of how the Agency will accelerate its regulatory procedures so that marketing authorisations of safe, effective and high-quality COVID-19 related medicines can be granted as soon as possible. The rapid procedures described in the inventory can accelerate every step of a medicine's regulatory pathway and the Agency is fully mobilised to deliver these fast-track assessments in the shortest possible timeframes while ensuring robust scientific opinions are reached.

"Supporting the development and marketing authorisation of safe, effective and high-quality therapeutics and vaccines as soon as possible is one of EMA's top priorities in the COVID-19 public health emergency. Together with our scientific committees and working parties, we have adapted our procedures in order to significantly shorten our own regulatory timelines for the review of new medicines and vaccines against COVID-19," said Executive Director Guido Rasi. "However, the rapid approval of therapeutics and vaccines will only be possible if applications are supported by robust and sound scientific evidence that allows EMA to conclude on a positive benefit-risk balance for these products."

These "rapid" procedures stem from EMA's emerging health threats plan. The flexible and fast review of medicines is supported by **EMA's pandemic Task Force (COVID-ETF)**, which brings together in one group the best scientific experts from the EU regulatory network. It will work closely with EMA's human medicines committee (CHMP) for optimal and fast coordination of activities related to the development, authorisation and safety monitoring of medicines and vaccines against COVID-19.

Accelerated support during research and development

For products under development, in early stages and/or before the submission of a marketing authorisation application, mechanisms put in place by EMA include:

- **Rapid scientific advice**, through which developers can receive prompt guidance and direction on the best methods and study designs to generate robust data on how well a medicine or vaccine works, how safe it is, as well as on the manufacturing and control process to establish its quality. In the context of COVID-19, fees for scientific advice are waived and the procedure is reduced to a maximum of **20 days**, compared to normally 40–70 days.
- **Rapid agreement of paediatric investigation plans (PIPs) and rapid compliance check**. The total review time for a PIP for COVID-19 products will be reduced to **20 days**, compared to normally up to 120 days active review time. In case needed, EMA also carries out a check to ensure companies comply with the agreed measures listed in each PIP before a marketing authorisation can be submitted, which will now also be reduced to 4 days.

All these accelerated mechanisms will require developers to submit well-prepared dossiers to EMA. The Agency therefore continues to encourage developers of vaccines or therapeutics against COVID-19 to make contact as soon as possible, to discuss their strategy for evidence-generation, by emailing 2019-ncov@ema.europa.eu. Depending on the maturity of the development, initial discussions on the various mechanisms to fast-track development and approval will take place, with priority given to the most relevant proposals.

Accelerated evaluation in authorisation and post-authorisation procedures

According to the EU pharmaceutical legislation, the standard timeline for the evaluation of a medicine is a maximum of 210 active days. However applications for marketing authorisation for COVID-19 products will be treated in an expedited manner:

- **Rolling review**. This procedure, used in a public health emergency, allows EMA to assess data for a promising medicine as they become available on a rolling basis. Under normal circumstances, all data supporting a marketing authorisation application must be submitted at the start of the evaluation procedure. In the case of a rolling review, CHMP rapporteurs are appointed whilst development is still ongoing and the Agency reviews data as they become available. Several rolling review cycles can be carried out during the evaluation of one product as data continue to emerge, with each cycle requiring around **two weeks**, depending on the amount of data to be assessed. Once the data package is considered complete, a developer submits a formal marketing authorisation application to EMA which is then processed under a shortened timetable.
- **Accelerated assessment**. This procedure can reduce the review time of products of major interest for public health from 210 days to less than 150 days. In practice, where there is an urgent public health need, assessment timelines will be reduced to the absolute minimum.
- EMA is ready to apply **further flexibility,** where it is established that shortening of any other procedural step could have an important public health impact in dealing with the COVID-19 pandemic.

The various rapid procedures are also available in the context of extensions of indications for already approved medicines, which are being repurposed in the fight against COVID-19.

The inventory also describes the support EMA can provide in the context of **compassionate use** programmes. Such programmes are set up at the level of individual EU Member States, to give patients access to treatments that are still under development and that have not yet received a marketing authorisation. EMA can provide scientific recommendations as to how these medicines should be used in this context, to support a harmonised EU-wide approach.

Regulatory Matters

Master trial documents for increased efficiency and scientific integrity

SECTION EDITOR

Jennifer Clemens, ELS
jennifer.clemens@merck.com

Creating and developing content for a programme of clinical trials involves a balancing act between leveraging efficiencies, maintaining programme-level consistency, and ensuring scientific integrity. One approach to managing such content is to utilise programme-specific document templates, but templates leave the door open to undesirable edits that do not serve to increase efficiency, consistency, or scientific integrity.

In fact, stylistic edits and changes that do not serve to change the actual meaning of the template content create confusion when comparing documentation across the programme in the hopes of understanding the key differences and associated rationale. Imagine, for example, a programme of five studies in which the table of contents for a given document type was aligned for four out of five of the studies, but for the fifth study, a key stakeholder decided that a given section should be moved to another location in the document or nested below a different heading as compared to the other four documents. Assuming the audience of the fifth document was familiar with the previous four studies, the reader of the fifth document would be perplexed as to why an entire section was removed from the fifth document and would waste time searching for the missing section and/or trying to understand why that section did not apply to that final study.

More seriously, changes to the study approach based simply on personal preference rather than on scientific justification can lead to a situation where data across studies in the programme are not easily comparable (i.e., apples to oranges rather than apples to apples). In both scenarios (i.e., stylistic and content edits), a consistent medical writer working across the programme of studies may serve as the gatekeeper and liaise between trial teams to align language and the approach as closely as possible, but this decreases efficiency and cannot completely prevent inevitable unnecessary differences in the content.

Following recent FDA guidance that defined the imaging charter document as "either a single document or a series of technical documents", it became possible to reorganise content into two separate documents.[1] Using the imaging charter as an example, evaluation of a given programme-level template revealed that 80% of the content was expected to remain consistent across studies in the programme and that only up to 20% of the content could reasonably be expected to vary due to individual study needs and differences (Figure 1).

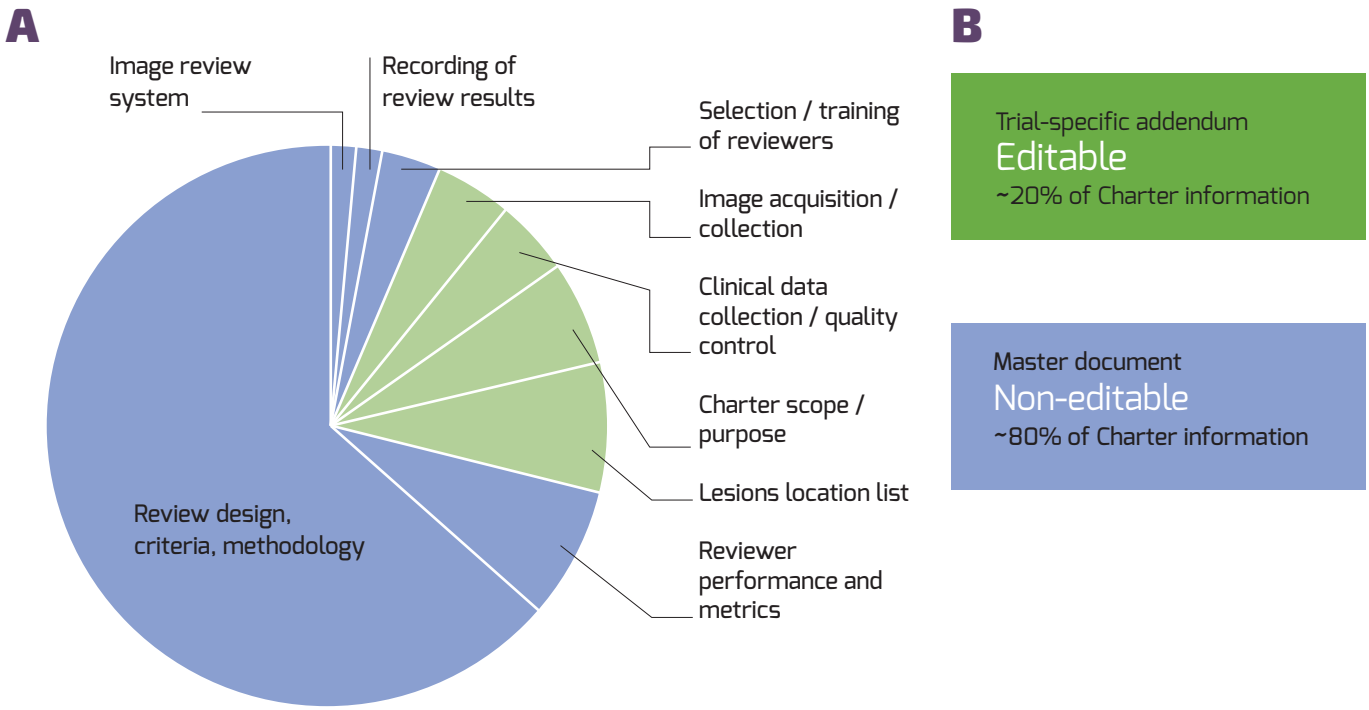


Figure 1. Anatomy of an imaging charter

(A) The pie chart represents a breakdown of protocol-specific charter document composition with respect to non-editable (blue) and editable content (green).

(B) The boxes represent the distribution of editable and non-editable content in a master charter.

Following this evaluation of the imaging charter template, the content could be reorganised into two documents. One root-level document, the master document, contains the locked content that should not change across a programme of studies; a second document includes only the protocol-specific clarifications and rationale for any changes from the programme-level approach, if applicable.

Application of this approach resulted in not only improved efficiency but also in increased consistency and scientific integrity, by discouraging any stylistic and unjustified trial-level alterations and teasing out the trial-specific information, thereby increasing transparency. Another added benefit was in the case of a required and justified programme-level change. Such a change could be applied once to the master document, thereby eliminating the need to make the same edit to each individual study document and also proactively removing the potential for additional edits to be made during document revision, which could potentially result in additional increased variance across the programme.

Splitting information across a locked root-level document and a second document that can be adapted for trial-specific information may not apply to every type of content. However, a modified approach that uses this same concept can be applied. For example, programme-level templates can be utilised but can be modified to include locked sections of content.

What is important to consider before applying a master document approach for a given trial document is whether the bulk of the content is specific to a given process, system, or programme, and is not expected to change significantly at the programme level. Once this is determined, the content can be organised in a new way to support improved efficiency, consistency, scientific integrity, and transparency.

Reference

1. U.S. Department of Health and Human Services, Food and Drug Administration. Clinical Trial Imaging Endpoint Process Standards – Guidance for Industry. April 2018 [cited 2020 Feb 29]. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/clinical-trial-imaging-endpoint-process-standards-guidance-industry

Kassel Fotinos-Hoyer, MA
Parexel International GmbH
Berlin, Germany
Kassel.Fotinos-Hoyer@parexel.com

Janina Wolf, PhD
Parexel International GmbH
Berlin, Germany
Janina.Wolf@parexel.com

Regulatory Public Disclosure

Editorial

Activities around clinical documents disclosure have been slow since September 2018 when I last published this section. As many of you receive emails from the CORE Reference website (sign up at https://www.core-reference.org/subscribe), you have been able to keep up with interim developments. This same information is regularly archived at: https://www.core-reference.org/news-summaries/ and https://www.emwa.org/sigs/regulatory-public-disclosure-sig/ and comes to you in the monthly EMWA Newsblasts, so you have been well supported via multiple open communication channels.

Broadly, the status quo remains… The three main regulators in the ICH family contributing to the disclosure narrative hold completely different positions at present:

A. EMA continues to hold clinical data publication activities (https://www.ema.europa.eu/en/human-regulatory/marketing-authorisation/clinical-data-publication/support-industry-clinical-data-publication). There is no indication of when activities may resume, or if public disclosure will apply retrospectively to documents, if or when activities resume. For this reason, we are best advised to maintain our awareness and continue to write our clinical documents in proactively anonymised fashion.
B. Health Canada is actively disclosing clinical documents (https://clinical-information.canada.ca/search/ci-rc) with guidance broadly similar to that of EMA (https://www.canada.ca/en/health-canada/services/drug-health-product-review-approval/profile-public-release-clinical-information-guidance/document.html), but with regulators discouraging redaction in favour of proactive authoring (qualitative anonymisation) and ultimately quantitative anonymisation methods
C. FDA is considering its options. After soliciting opinion on how FDA might best support disclosure of clinical documents (https://www.regulations.gov/docketBrowser?rpp=25&so=DESC&sb=commentDueDate&po=0&dct=PS&D=FDA-2019-N-2012) and announcing the conclusion of its Clinical Data Summary Pilot in March 2020 (https://www.fda.gov/news-events/press-announcements/fda-continues-support-transparency-and-collaboration-drug-approval-process-clinical-data-summary?utm_campaign=032620_PR_FDA%20Supports%20Collaboration%20as%20Data%20Summary%20Pilot%20Concludes&utm_medium=email&utm_source=Eloqua), FDA is not currently disclosing clinical documents but has identified a possible approach for disclosing study reports, the framework of which includes the following principles:
    - A centralised international library managed by an independent body would be set up where information is made available to the public, rather than each regulatory authority having its own system
    - An on-demand system would be set up where some documents, e.g., clinical summaries, index of study reports, would be automatically published. The public could request documents and the sponsors would add them to the library
    - Anonymisation and disclosure standards would apply; PHUSE standards are particularly mentioned (https://www.phusewiki.org/wiki/index.php?title=Data_ Transparency)
    - Sponsor commitment to use the international library system would be voluntary.

The trend of the pharmaceutical industry being better at posting summary clinical trial results to public registries than other sections of the clinical trial community continues (https://www.thelancet.com/pdfs/journals/lancet/PIIS0140-6736(19)33220-9.pdf).

Art Gertel (CORE Reference Strategist) and I had planned to present this topic at the cancelled EMWA Conference in Prague planned for May 2020. Due to its time-sensitive nature we have made our slide deck available as an educational resource at: https://www.core-reference.org/publications/

SECTION EDITOR

Sam Hamilton
sam@samhamiltonmwservices.co.uk

Other news in brief

CORE Reference

In August 2019, the CORE Reference development team (Budapest Working Group, BWG) published a paper titled: **Critical review of the TransCelerate Template for clinical study reports (CSRs) and publication of Version 2 of the CORE Reference (Clarity and Openness in Reporting: E3-based) Terminology Table** (http://dx.doi.org/ 10.1186/s41073 - 019-0075-5). Our paper includes a detailed assessment of TransCelerate's November 18 CSR template in the form of an "additional file" comprising a 44-page replica of their template marked up with our 69 consolidated comments.

In December 2019, TransCelerate released updated versions of their CSR template and SAP template (referred to as "assets"). These new assets and supporting resources reside at a **new page location** (https://transceleratebiopharmainc.com/assets/clinical-content-reuse-assets/). This relocation of assets has taken place since the publication of our paper. To download the assets, you need to complete an online form.

The December 2019 TransCelerate CSR template is supported by a slide deck titled "Summary of Changes in 2019 Release". This 40-slide deck includes a rationale for each change. From Slide 25 or thereabouts, the rationale for change frequently includes "Feedback from CORE review" or "CORE feedback". TransCelerate notes that this new release brings their template into "alignment with CORE"; however, there are no specifics provided as to how comprehensively the CORE feedback was addressed and incorporated.

TransCelerate have not made any contact with the BWG. The BWG have not reviewed the December 2019 TransCelerate CSR template.

Art Gertel and I had planned to present this topic in May 2020 at the EMWA Conference in Prague that was cancelled. Due to its time-sensitive nature we have made our slide deck available as an educational resource at: https://www.core-reference.org/ publications/

Resources

Two excellent white papers:

- https://cdn2.hubspot.net/hubfs/200783/PC20257%20Clinical%20Data%20Disclosure%202020/Managing%20Privacy%20Risk%20in%20Data%20and%20Document%20Sharing.pdf
  This is a white paper from experts at Astrazeneca and d-wise gets to the nub of what medical writers need to understand around proactive anonymisation of data and documents, and is a great summary of what many of us have been discussing for some years now. In their own words, the authors address: "How should sponsors manage data they share considering what's already been shared? What techniques exist to support sponsors in navigating the reality of human error and the limits of technology?" There are some good screenshots of anonymised and redacted data and documents towards the end of the document.
- https://www.d-wise.com/white-papers/preparing-clinical-study-reports-for-external-sharing?utm_campaign=%5Bobject%20Object%5D%20Transparency-CBI&utm_source=hs_email&utm_medium=email&utm_content=82865504&_hsenc=p2ANqtz—v6Sex-ip2pz9n8murNSD4_pfwcYNoTvViOr EUzKWYSEojKl6EwrPW2P1WKlBbueX4B8gdBhr8x7OdL_UlOPG777QKyNIj49Xf5lY-pyS1aFX2rfo&_hsmi =82865504
  At a recent CBI Clinical Data Disclosure, Transparency & Plain Language Summaries event: "Sharing to Power Innovation", Cathal Gallagher (EMA TAG member) outlined the necessary steps for internal and external sharing in his presentation and white paper, "Preparing CSRs for external sharing". This excellent summary gets to the nub of why industry need to better support medical writers with CSR proactive anonymisation. Also read Cathal's interview with me (on page 58).

Data transparency workshop

On November 11, 2019, a data transparency workshop with EMA was held in Amsterdam as part of the PHUSE EU Connect 2019 event. The event was led by Jean-Marc Ferran of the PHUSE Data Transparency Working Group and EMA representative Anne-Sophie Henry-Eude. They were joined by several TAG members to address questions during the Q&A panel session.

Key topics discussed during the workshop included:

- EMA Policy 0070 Phase 1 and handling of the backlog
- EMA Policy 0070 Phase 2
- EMA – Health Canada Collaboration in data transparency.

Jean-Marc provided a summary of the workshop to the PHUSE community via a webinar on November 20, 2019. View the recording at https://www.youtube.com/watch?v=eQGyL 4SI1K0 (approximately 11 to 25 minutes).

The slides are available here: https://www.phusewiki.org/docs/Working Groups/Webinar/November%202019/EUCon19%20-%20DT%20 Workshop%20-%20Webinar%20Slides%20 -%20v000b.pdf

Recent events impacting transparency and disclosure

EUDAMED delay and the impact on devices transparency

In an article for the Regulatory Affairs Professional Society, Raquel Billiones reviews possible ramifications caused by the delay in launching the European Union's new electronic database, the European Database on Medical Devices (EUDAMED). The article is available at https://www.raps.org/ news-and-articles/news-articles/2020/4/ eudameds-delay-what-happens-to-transparency-for-cl.

COVID-19 impact on clinical trial disclosure and transparency

Regulatory authorities have released guidance documents focusing on the impacts of the COVID-19 pandemic on study start-up activities, changes to ongoing study procedures, and items considered urgent safety matters during this pandemic. This PHUSE blog (https://www.phuse.eu/blog/the-impacts-of-covid-19-on-clinical-trial-transparency-and-document-disclosure-phuse-ctt-project) considers the impacts of COVID-19 on clinical study disclosure and transparency, offering guidance from industry experts on what may require immediate action, as well as consideration of future implications

Health Canada issues notices of nonconformance

In recent months, HC placed identical notices on submissions packages from Lilly, Novartis, Seattle Genetics and Gilead which state that in respect of CSR narratives there are "…extensive redactions to the patient information... redactions do not conform to HC guidance which encourages... other transformation methods…"

Read the full Lilly notice here as an example: https://clinical-information.canada.ca/ci-rc-vu.pdf?file=m1/ca/HC%20STATEMENT%20REDACTED%20PATIENT%20INFORMAOIN%20ENFR.pdf&id=128554

Without a change from retrospective redaction to the proactively anonymised authoring of CSR narratives that is actively encouraged in CORE Reference, we can surely expect to see similar notices on future submissions.

Medical Devices

SECTION EDITOR

Kelly Goodwin Burri
kellygoodwinburri@stryker.com

Editorial

May 2020 was set to be a busy month for medical device writers in Europe. The EU Medical Device Regulation (MDR) was to take effect on May 26, and we were anticipating lively discussions on the final MDR preparations with our fellow writers at the 50th EMWA conference. For this issue, Cherry Malonzo had originally intended to provide you with a recap of the Medical Device Expert Seminar Series (ESS) from the conference in Prague. That is until the coronavirus pandemic arrived in Europe, the EMWA conference was cancelled, and even the implementation of the MDR was postponed. With no ESS to report on, Cherry proposed to reach out to our medical writing colleagues to hear how they are coping with this unprecedented situation, both personally and professionally. In this issue, we can share with you three of the responses from colleagues weathering the COVID-19 lockdowns in England, Germany, and Switzerland. I hope you are all staying safe, healthy, and sane in these challenging times.

Kelly

Medical writers under lockdown

When the coronavirus lockdowns were put in place across the globe in a span of a few short weeks, it felt as though we had all been thrown into the set of a bad movie. Suddenly, we all found ourselves re-orienting our entire existence to the new normal, learning to find the best means to cope with the circumstances (and with each other). Regardless of location, it is not hard to relate to the following accounts and reflections of our colleagues Jane, Claudia, and Payal.

Cherry Malonzo Marty
cherrymalonzo@m3writing.com

Jane Edwards, England

I have two types of a rare autoimmune disease, vasculitis, which means that my immune system attacks my blood vessels and causes inflammation. The first type is granulomatosis with polyangiitis (GPA), which affects my small blood vessels, leaving lasting damage to my kidneys, eyes, nose, and ears. The second type is large vessel vasculitis, Takayasu Arteritis (TAK), which has caused inflammation in the main aortic root and arch.

The COVID-19 pandemic has created a crazy time, and it is extremely difficult for those of us with underlying health issues. I am used to being isolated after 8 years of being ill, but I now find myself surrounded by people 24 hours a day. Having an energised husband, an emotional teen, and a very bouncy dancing 9-year-old in the house is leaving me exhausted. But I am also glad to have them all close and to be able to do fun things together. We have been shooting netball hoops in the garden, playing Jenga, and doing a YouTube Joe Wicks PE podcast some mornings.

The whole of the UK is on lockdown, and people should not leave their homes except for essential shopping, daily exercise, and critical medical appointments. We have a list of key workers, including medical staff, who are also allowed to travel to and from work. I work for BSI Notified Body as the Global Head of Communications, and a significant number of the team were already partly working from home, so the adjustment has been gentle.

Because of my level of prednisolone and the fact that I have Rituximab infusions every 4 months, I am considered high-risk and advised to go into "shield" mode by the UK government. We are advised not to leave the house for any reason, and to enact "social distancing" inside the house with our family. My whole family has been social distancing for about 20 days. The advice is to stay 3 feet from the family, but because we have been cut off, we have decided that we will not do this, partly because we live in a quiet area and can avoid people. I am not advising anybody else about what they should do, we must look at our own situation.

I have found it beneficial to just stop using all

social media. I couldn't cope with all the comments from people, whether they were claiming to be experts, making light of the situation, or having a dig at those who were taking it seriously or conversely ignoring the warnings. This has helped me to almost ignore the situation. I have stopped watching the news, except for listening to the government briefing just once a day. I believe the hard part of this situation is to remain positive and try to keep a routine, both for work needs and family time. Maintaining the balance will be how we survive this difficult time.

Jane Edwards
Jane.Edwards@bsigroup.com

Claudia Frumento, Germany

To be honest, being a freelance medical writer who works from home, I was not expecting my life to change much under the self-imposed coronavirus quarantine. I wake up at around 7 am, enjoy breakfast, read the newspaper, start working in my pyjamas, and at some point, when my dog (Greco) gets nervous, I get dressed and go for our daily walk in the woods.

But after only 2 weeks, I realised there were changes. There were some changes that I might have expected first in a few years, with retirement: my husband is also at home now, every single hour of the day, every single day and night! It is not that I don't love him, but he uses MY phone (ringing all the time), he uses MY office, and he uses MY printer. By the way, my husband is working very closely with the health authorities of the country and he is ALWAYS in conference calls using MY phone line, and MY calls are diverted or are lost in the corona ether.

Our poor dog Greco does not understand why he can't doze any more in his preferred corner of the office and bark once in a while when a rabbit comes into the garden. Before coronavirus he used to bark at the footballs that came in flying from the playground. Now he gets all confused when he finds me working downstairs in the living room.

But the real problems start around dinner time when my husband finally comes downstairs. The new "corona news" discussed during the endless conference calls are described in detail over dinner. If I had managed to forget about it during the day, it all comes over me again. The fear, the "German angst", what this virus is going to do to us all, why we are part of that high-risk group that is dying like flies in other countries, why men seem to die more than women, why this will have a tremendous negative impact in the economy! The only thing that has not yet been discussed at high health authority levels is why Germans and most Europeans seem to have a fixation on toilet paper in times of crisis!

Yesterday, I issued an ultimatum to my husband: think of something nice to discuss during dinner, simply one topic that is not related to corona. He answered with a smile: how about the Netflix series *The Crown** or the *Walking Dead*?

*corona = crown in Spanish

Claudia Frumento
claudia.frumento@t-online.de

Payal Bhatia, Switzerland

It was a Sunday morning when I heard from my employer that I would be working from home. With COVID-19 picking up pace, this arrangement was a big relief. But was it? Several questions started sprouting up in my mind: Was it too late? How bad will things get? How long will the lockdown last? Should we stock up groceries? What if the daycare is closed? Even if daycare is not closed, should we continue sending our 4-year-old? And on and on I went.

Now, almost 4 weeks later, I can say that it was fear. Not the fear of dying from the virus, but the fear of the unknown. The fear coming from the uncertainty that surrounded (and still does) the circumstances and our ability (or inability) to deal with them. The fear of not making the right decisions and the fear of losing it all with the lockdown. But our doubts have started to settle, and we are managing to sail through, just like everyone around the world is. We continue to send our daughter to daycare, and my husband and I have found peace by dividing our home into separate office zones. I took the office room, not because I am mean, but because I work 2 days from home in my 4-day work week and it just seems easier to continue with the status quo. He has set up his office in the dining room. We communicate more than usual to ensure we do not have overlapping meetings but schedule our lunch together. Sharing domestic work is no longer relegated only to the weekends.

Speaking of weekends, keeping our child entertained is a challenge. My go-to activities so far have mainly included teaching my daughter to write, inventing silly stories and songs almost the entire day, letting her tune into audio stories every now and then, reading to her even more, taking walks in the barren forest in the neighbourhood, hosting imaginary birthday parties (and dressing up for them), cleaning out closets and cabinets, baking, and preparing our terrace for spring. My sanity comes from organisation around the house, good nutrition, mindfulness, reading, and keeping myself well-informed – but not overinformed. Besides that, I am staying hydrated, using the stairs at least once a day, and trying hard to practice minimalism in everyday life. By staying home, we are all doing our part. Let's hope it's enough!

Payal Bhatia
payal.bhatia@medicalminds.ch

EU MDR implementation postponed

The European Commission has adopted a proposal to delay implementation of the EU MDR by 1 year, to May 26, 2021.[1,2] The delay is intended to ensure essential medical devices remain available and allow medical device manufacturers to prioritise efforts to fight the COVID-19 pandemic. While the MDR date of application will move to 2021, the planned implementation of the in vitro Diagnostics Regulation (IVDR) remains unchanged with application from May 26, 2022. The proposal received the full support of the European Parliament and the Council needed through an accelerated co-decision procedure to become effective before the original May 26, 2020, MDR implementation date.

References

1. European Commission. Commission postpones application of the Medical Devices Regulation to prioritise the fight against coronavirus. 2020 [cited 2020 Apr 6]. Available from https://ec.europa.eu/commission/presscorner/detail/en/ip_20_589.
2. European Commission. Guidance on medical devices, active implantable medical devices and in vitro diagnostic medical devices in the COVID-19 context. 2020 [cited 2020 Apr 6]. Available from https://ec.europa.eu/docsroom/documents/40607.

Kelly Goodwin Burri
kelly.goodwinburri@stryker.com

Getting Your Foot in the Door

Editorial

Bringing you some ❤️💖 in the time of corona.

Now, more than ever, do we need online resources for training, re-training, and cross-training. Thanks to Diana Ribeiro for sharing the information below to grow our GYFD online resources:

- **Writing in the Sciences from Coursera** is free without the certificate and touches on several basics.
- **The Health Writer Hub website (healthwriterhub.com)** run by Michelle Guillemard, is aimed at health writers but it has a lot of useful information for medical writers, too. The blog is free, and there is also a free email course if you subscribe to her newsletter. Michelle also offers courses for a fee, such as "Introduction to Health Writing" and "How to Become a Freelance Health Writer".
- **Stgilesmedical(https://www.stgmed.com/events)** offers several e-learning modules on medical communications for a fee.

SECTION EDITOR

Raquel Billiones
medicalwriting@billiones.biz

Lots of thanks also go to Namrata and Gauri for sharing their stories of resilience and determination below. Just what we need, especially when job hunting in times of crisis.

To the medical writing community, stay strong, safe, and healthy.

Raquel Billiones

Big feet make for a great door stopper

Getting Your Foot in the Door – this title transports me back to a moment when I was a little girl trying on new shoes and being told, for the hundredth time, what big feet I have – pun intended.

Let me start this article by stating, upfront, my mantra for getting one's foot in the door: **Be a stubborn optimist.**

My professional background is in dental medicine. I studied and practised dental surgery in India and eventually settled in Switzerland post marriage. The questions I had initially were straightforward and simple – first, what is fondue and how do I eat it? Second, is a crab-like, sideways walk spanning 2 hours normal for my first ski lesson? Third, and most importantly, how do I practice dentistry again? Unfortunately, the only success story I have here is figuring out how fondue consumption works.

I found myself, after much struggle, at the brink of an inevitable career change. As I contemplated the options that best suited me, I started writing two blogs, one about my experiences as an ex-pat and the other dedicated to dentistry. I realised that writing has been an old companion, and in today's day and age, something where I could find good value. It was sometime in 2017 when I googled the words "medical writing" for the first time. That year, I was introduced to two new bodies – my precious son and EMWA.

EMWA's free online resources for new medical writers introduced me to the different types of medical writing. When confronted with a choice between medical communications and regulatory writing, a quick job search led me to quantify the latter as being more in demand. However, the job description of a regulatory writer was akin to High Valyrian,[1] which brought me back to medical communication.

I got busy building my new skillset. Editing, proofreading, understanding the scientific methodology and what makes a good research paper. The thought process behind a research article as an author was new to me as I wasn't trained in this skill as a medical doctor. I had referred to many research articles to guide my clinical decisions but had never written a research paper or a thesis.

> I did killer tooth extractions and root canal treatments but had never written a thesis before. Pardon me, I retract my words – "killer" and "extractions" should never be mentioned in the same sentence.
>
> *Disclaimer: Any resulting nightmares of dental clinics are not the responsibility of the author.*

I found myself having different transferrable skills than my competition who were (and still are) mostly PhD professionals. In fact, there was a good chance that my potential hiring manager would be a PhD graduate too. Consequently, I focused on the transferrable skills that PhDs

[1] The Valyrian languages are a fictional language family in the fantasy novels by George R. R. Martin and in their television adaptation *Game of Thrones*. High Valyrian is famously spoken by *Daenerys Targaryen, The First of her Name, Queen of the Unburnt, Khaleesi of the Great Grass Sea, Mother of Dragons, Protector of the Realm, Queen of…*" (please read the novels for the rest of it).

bring, which led me to Cheeky Scientist. This proved quite useful as they prepare PhD professionals for entering the industry. I unashamedly declare that I was a silent weed to their entire business model. Thank you, *Cheeky Scientist*.

EMWA also introduced me to medical writing in the medical device industry. As comfortably as I had used medical devices as a clinician, I also had knowledge of drugs and their development. I found myself briefly studying both Pharma and MedTech industries to judge what suits me best. Quickly I realised that the High Valyrian only gets higher. The number of abbreviations used is mind-boggling. CSR, CER, ISO 10993, MDR, CSP, clinical trials in Pharma = clinical investigations in MedTech and so on. To end the confusion, I took to studying some more.

I did most online courses available on Coursera and Udemy (see references) to get accustomed to the new jargon. Additionally, I chose five varied job adverts as templates for my new self-made syllabus, deciphering the lingo and learning what I could to build my pitch. I started updating myself on current affairs in the medical device industry as this industry was more in tune with my professional skills. Learning about MDR helped me to foresee a great need for medical writers in the coming months. Armed with this new knowledge, I started applying for medical writing positions, confident that my new and refined CV should get me noticed.

I received a great number of rejection emails during that time. By mid-2019, the solution was clear – I must step out and be seen; otherwise, this Catch 22 situation will become a Catch 72[2] with wrinkles and a jarring pain in the hip if I don't do something about it. So I decided to attend my first EMWA conference in Vienna in May 2019.

> Rejection – the more you face it, the more you train your mind to not have an emotional response to it every time it repeats.

Armed with my fancy business cards, I made the conference my practice ground, polishing my newborn pitch and amending it as per to the person I spoke to. Each person I met was a mentor to me. Their smiles and openness validated my choice to attend it. I only managed to enrol in one workshop but quickly discovered its advantages. I was absorbing every mistake I made during my conversations with subject matter experts and correcting it on the go. The EMWA conference in Vienna helped me rekindle something that had been lost – my self-confidence.

Be a stubborn optimist!

Being an optimist is crucial to a career change. The journey to attain this optimism is full of rejections, self-doubt, and monotony that eventually peaks and propels the person to undergo drastic changes to become an individual who deserves better. I found that shedding my title of Doctor and addressing head-on what I lack was mandatory. The stubbornness in this optimism is patiently self-constructed as a result of hard-hitting circumstances that everybody must face a few times during the course of a single lifetime. A resulting realisation that we are all temporary and sharing our time on this planet leads to an epiphany that there is no room for negativity or judgement as I might leave as quickly as I have arrived. If I must be stubborn, let it be for optimism.

> Pessimism is a waste of time – literally.

As a result of all of these collective experiences, I managed to get my oddly sized feet through the door and landed my first job as a project associate in clinical affairs/regulatory affairs and quality assurance at a medical device consulting firm. Over the course of 6 months, I've written multiple clinical evaluation reports (CERs), attended internal and external audits and forged wonderful lasting relationships with my new colleagues. I'm now exposed to a new language style of corporate communication which is quite different from what I was accustomed to as a medical professional. I am under construction every single day, exhilarated by all the new knowledge and look forward to my next new challenge.

> Big feet may take longer to get through the door, but they also make for a great door stopper.

After years of struggle, I can confidently say that I'm the doctor who can write CERs and more thanks to my new job experience and EMWA's support. I look forward to keeping the wheels turning, helping all those who currently find themselves at the brink of a career change into medical writing.

I raise my glass to the stubborn optimists.

Resources:

1. Useful courses on Coursera:
   - Introduction to Systemic Review and Meta-Analysis – https://www.coursera.org/learn/systematic-review
   - Pharmaceutical and Medical Device Innovations – https://www.coursera.org/learn/pharma-medical-device-innovations,
   - Drug Development - https://www.coursera.org/learn/drug-development
   - Design and Interpretation of Clinical Trials - https://www.coursera.org/learn/clinical-trials
2. Useful courses on Udemy:
   - ISO13485:2016 – Design and Development of Medical Devices – https://www.udemy.com/course/iso-134852016-design-and-development-of-medical-devices/
   - Applied ISO14971 Medical Device Risk Management – https://www.udemy.com/course/applied-iso-14971-medical-device-risk-management/
3. Medical Device Made Easy podcast series by Monir El Azzouzi – https://podcast.easymedicaldevice.com/
4. British Standards Institution (BSI) white papers – https://www.bsigroup.com/en-GB/medical-devices/resources/whitepapers/
5. Greenlight Guru – https://www.greenlight.guru/
6. Author's websites – http://dentalyoda.blog, – https://diaryofadiscoverer.wordpress.com

Namrata Upadhyay
dr.namratau87@gmail.com

[2] Catch 72 (*noun*) – a dilemma or difficult circumstance from which there is no escape… for people in their 70s.

My journey into the industry

A few *Medical Writing* issues ago I wrote about how I started medical writing as a freelancer. After a freelance stint, predominantly in the in vitro diagnostics (IVD) industry, I decided to switch to an employed position. The question was how? In this article, I am sharing my journey into the pharmaceutical industry.

The most challenging part of being a freelancer is to gain/maintain customers while expanding your professional skills. Having decided to change my career trajectory, I found myself running around in a circle of regulatory writing for pharma or medical devices, medical communications, promotional writing! What should I focus on? The choices were endless. The more I delved into the realm of medical writing, the more I realised that I cannot possibly stretch myself in all directions. Therefore, I needed to streamline my efforts in finding a balance between focusing on topics along my freelance path (that could lead towards future customers) and following my interests.

Looking back, there were four important steps that connect my journey like a thread.

Belonging to a professional society

Professional societies provide insights into the requirements for each career. Fortunately, I was already a part of EMWA and cannot stress enough how invaluable EMWA is for my medical writing career. The workshops offered by EMWA's professional development programme were the perfect platform to implement my strategy. With every workshop I took, my perspective and interest changed. I found out what I do like and what I would rather let go of. Moreover, it was very encouraging to hear from Gillian Pritchard and Raquel Billiones that regulatory medical writers can switch between writing for the pharma and medical device industries once they know the basics. I no longer felt that my IVD background did not matter. On the contrary, it was an opportunity. For example, the use of companion diagnostics in personalised medicine would require co-development of the review and approval process and the new MDR and IVDR regulations would bring the two industries closer to the pharmaceutical regulatory pathway.

EMWA also provided me with an opportunity to network. This was the next essence that paved my journey.

> "The small but powerful word 'YET' was enough for me to realise that I do possess the required transferable skills as a PhD. It was just a matter of time until I would get experience under the right circumstances."

Networking

EMWA conferences are a great venue not only to learn new tools and tricks of the trade but also to interact with potential clients and other members that are active in the field. Networking and discussing your point of view not only lets you identify your qualifications but also makes you aware of the challenges that others are facing. Sometimes, the exchange of ideas leads to solutions, such as collaboration amongst freelance colleagues. At other times, it leads to unexpected and much-needed advice. Beatrix Doer's words at the Warsaw conference still ring in my ears. When I told her about the almost standard response of "you are not experienced" that I got during my job application process, she added a small three-letter word "YET" after the word "experience". She was right and it changed my perspective completely! The small but powerful word was enough for me to realise that I do possess the required transferable skills as a PhD. It was just a matter of time until I would get experience under the right circumstances. So, I continued with my efforts of investing in myself.

Spruce up your CV

For a lot of people with an academic background, the question of "what do I put in my CV" is an uncomfortable issue. Suddenly, the imposter syndrome sets in and you freeze. This is bound to happen if you think that your entire future professional career depends on your CV. While it is true to some extent, I have learnt to look at the CV from a different perspective. I let the CV help me in truly understanding my qualifications. Only when you understand your qualifications can you articulate them in a way that would be attractive to a potential employer. In other words, you need to "market yourself". There is a trove of CV tailoring websites/software out there that will give you a nice final document. However, it is the content and the way it is presented that matters. Unless you have thoroughly understood your qualifications, you will always have a hard time convincing the potential employer. A good exercise that helped me was making an Infographic CV, which was introduced to me by EMWA colleague Carola Krause (who over time has also been a great source of advice) during a seminar. Essentially, an Infographic CV contains concise information in a graphic form, but with a bird's eye view. It was quite challenging to weed out the real information that I wanted to emphasise on and leave non-relevant experience out. After the bird's eye view of my CV, it was easier to tailor the content for job-specific requirements. Subsequently, I did create a standard CV, which was more detailed.

Contribute

It is quite important for medical writers to publish articles, not only to sharpen your writing skills but also to realise that your content has an essence that convinces the reader. The topics can range from more general ones, for example, your career path or more specific ones, for example, changes in the regulatory guidelines. EMWA's journal *Medical Writing* is an excellent platform to share your experiences and know-how. It was very satisfying to be able to give advice when people reached out to me after reading my story as a freelancer. Inspired by this experience, I volunteered to write this article. There are several other open-source platforms or blogs via which you can give back to the medical writing community.

At this stage in my career as a regulatory medical writer, I feel more confident in defining my professional identity. Moreover, I am excited to be part of an industry that strives to bring medical solutions to patients.

Gauri Jawdekar-Abraham, PhD
jawdekar.abraham.g@gmail.com

Journal Watch

Journal Watch is based on the French language blog, *Rédaction Médicale et Scientifique*, by Hervé Maisonneuve, available at www.redactionmedicale.fr.

SECTION EDITOR

Hervé Maisonneuve
herve@h2mw.eu

COVID-19 pandemic articles in journals: Lessons for the future?

The SARS-CoV-2 pandemic responsible for COVID-19 is testing the resilience and limits of our established system of science publishing. As a result, scientific journals may have to reevaluate their existing standards, roles, and economic models once the crisis is over. It is too early to predict what changes will endure, but I'm already certain of one thing: The first four months of 2020 have already altered the way scientific journals work. Here are some examples and observations about these new developments:

- Clinicians have run out of time to write up their results because the demand for medical personnel is so great. Other researchers, however, have found themselves homebound with lots of time to analyse old data and prepare it for publication. Journal editors and reviewers have not always been available to perform their usual editorial and vetting tasks.
- There has been great competition among prestigious journals: *Nature* (UK) and *Science* (USA) have raced to attract research papers; *The Lancet*, The *BMJ*, the *New England Journal of Medicine*, and *JAMA* are competing to attract medical research. Most of them have sought out "hot" papers, preferring these to others of lesser interest in a time of pandemic crisis. We have to ask ourselves: Did journals sometimes lower their standards in order to accept papers on hot topics?
- Most prestigious journals and publishers have created a site dedicated to COVID-19 articles. Nearly 100 academic journals, societies, institutes, and companies have signed an agreement to make research and data on COVID-19 freely available, at least for the duration of the outbreak.[1] Some journals waived their usual article processing charges. Some editing companies have offered to edit papers for free.
- The volume of published papers on COVID-19 is high: more than 16,000 articles were published in peer-reviewed journals between January and May 2000 (https://pubmed.ncbi.nlm.nih.gov/?term=covid-19). To meet the demands of this accelerated pace, a fast-track for peer review was used for most, and journals often resorted to published calls to find reviewers.
- The *NEJM* has reported receiving 40 COVID-19 papers per day, and accepting 2%.[2] *JAMA* published an editorial on the lapse in ethical standards of scientific reporting: "The editors have become aware that some of the patients described in some of these manuscripts, sometimes with overlapping authorship, have been reported in more than 1 submission".[3] Case reports based on the same sets of patients have been published in different journals.
- The quality of many of the published papers was poor, and at least 50% were deemed of little scientific interest.[3] High impact journals have published observational studies based on fewer than 10 cases, with poor case reports, and open, non-comparative non-randomised trials with fewer than 50 patients. Specialty journals have received papers rejected from prestigious journals.
- Chinese authors have been numerous, and their papers – in contrast to those published during the previous coronavirus pandemics – were only signed by Chinese authors. This change in authorship practice is a new development in scientific communication and needs to be evaluated after the end of the pandemic.
- Many new databases and websites have been created to compile the literature on COVID-19; the site of the Evidence for Policy and Practice Information and Coordinating Centre, UK, regularly updates the literature.[4] On April 1, 2020, this site listed 2,340 papers; some were excluded: 1404 (not primary data), and 169 (concerning other viruses); the other papers were classified as case reports (189), transmission/risk/prevalence (159), health impacts (143), diagnosis (95), genetics/ biology (80), case study/ organisation (72), treatment drugs (41), mental health

impacts (10), social/economic impacts (8), vaccine development (5), intervention/outcomes study (5).

- Preprints have gained enthusiastic support, even though before this epidemic, some authors and writers were resistant to their use; however, the number of COVID-19 preprints was difficult to estimate due to the great number of archives involved. Nonetheless, at least 2000 preprints related to COVID-19 were deposited between January and March 2020; for example, on April 6, 2020, medRxiv had 924 preprints, while bioRxiv had 279 preprints (https://connect.medrxiv.org/relate/content/181). For bioRxiv, 30% of these preprints remain unpublished, yet the majority are already posted onto bioRxiv close to or after submission.[5] We don't know if this observation will also apply to medRxiv COVID-19 preprints.
- The *International Journal of Antimicrobial Agents* published a series of poor papers from Didier Raoult and his team on the use of chloroquine to treat infection by coronavirus; one of them reported encouraging results with 19 patients but also revealed numerous biases.[6,7] The main objective was probably to be mentioned by the media, and indeed, it did get a US presidential tweet; the journal's editor and another editorial board member were co-authors of these papers; most papers were accepted with an expedited peer review of 12 to 24 hours. Exceptionally, ISAC (International Society of Antimicrobial Chemotherapy), owner of the journal, issued a press release with the following statement: *ISAC shares the concerns regarding the above article published recently in the International Journal of Antimicrobial Agents (IJAA). The ISAC Board believes the article does not meet the Society's expected standard, especially relating to the lack of better explanations of the inclusion criteria and the triage of patients to ensure patient safety.*

In the single month of March 2020, Didier Raoult and Jean-Marc Rolain (Editor in Chief) co-authored seven papers on COVID-19 in this journal.

Looking over these observations, we must ask ourselves: How will journals get back on track after this article pandemic? Will they re-install article processing charges and paywalls for the COVID-19 papers at some point? Will preprints become more accepted by the clinicians and researchers? Will journals change their peer review process so that fast-tracking and open reviewing become permanent features? Will they all switch to an open-access model?

Clearly, the COVID-19 pandemic has already had a dramatic impact on our daily lives and health. Continued monitoring will be necessary to assess whether – and to what extent – it will also alter the course of established processes for scientific publication.

References

1. Yasinski E. Journals open access to coronavirus resources. The Scientist. 2020 Feb 13.
2. Jarvis C. Journals, Peer Reviewers Cope with Surge in COVID-19 Publications. The Scientist. 2020 Mar 17.
3. Bauchner H, Golub RM, Zylke J. Possible reporting of the same patients with COVID-10 in different reports. JAMA. 2020 Mar 16. Available at: https://doi.org/10.1001/ jama.2020.3980
4. EPPI Centre. COVID-19: a living systematic map of the evidence. [cited 2020 April 3]. Available at: http://eppi.ioe.ac.uk/cms/Projects/DepartmentofHealthandSocialCare/Publishedreviews/COVID-19Livingsystematicmapoftheevidence/tabid/3765/Default.aspx
5. Anderson K. bioRxiv: trends and analysis of five years preprints. Learn Publ. 2019;33(2):104–9.
6. Gautret P, Lagier JC, Parola P, Hoang VT, Meddeb L, Mailhe M, et al. Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial. Int J Antimicrob Agents. 2020 Mar;105949.
7. Bik E. Thoughts on the Gautret et al. paper about Hydroxychloroquine and Azithromycin treatment of COVID-19 infections. Science Integrity Digest. 2020 March 24. Available at: https://scienceintegritydigest.com/

"Registered Reports" associated with increased proportion of negative results in the published literature

"Registered Reports" is a publishing format that emphasises the importance of the research question and the quality of methodology by conducting peer review prior to data collection. High-quality protocols are then provisionally accepted for publication if the authors follow through with the registered methodology. This format eliminates a variety of questionable research practices, including low statistical power, selective reporting of results, and publication bias, while allowing complete flexibility to report serendipitous findings. Currently, 242 journals use the Registered Reports publishing format (https://cos.io/rr/?_ga=2.48543974.1956374534.1585861906-633746582.1578172282).

Peer review occurs prior to observing the outcomes of the research. Manuscripts that survive pre-study peer review receive an in-principle acceptance that will not be revoked based on the outcomes, but only on failings of quality assurance, following through on the registered protocol, or unresolvable problems in reporting clarity or style.

A comparison of articles between standard reports and Registered Reports was made and published as a preprint (not yet published in a peer-reviewed journal).[1] I copied extracts from the Abstract:

> *We compared the results in the full population of published Registered Reports in Psychology (N = 71 as of November 2018) with a random sample of hypothesis-testing studies from the standard literature (N = 152) by searching 633 journals…*

> *Analysing the first hypothesis reported in each paper, we found 96% positive results in standard reports, but only 44% positive results in Registered Reports. The difference remained nearly as large when direct replications were excluded from the analysis (96% vs 50% positive results). This large gap suggests that psychologists under-report negative results to an extent that threatens cumulative science.*

Reference

1. Scheel AM, Schijen M, Lakens D. An excess of positive results: Comparing the standard psychology literature with Registered Reports. Preprint. Posted online 2020 Feb 5. PsyArXiv. Available at: https://doi.org/10.31234/osf.io/p6e9c

Mark your calendar

September 2021 International Congress on Peer Review and Scientific Publication (Chicago)

In 2020, consider performing some kind of research on peer review, with the objective of presenting a poster or a communication at the ninth International Congress on Peer Review and Scientific Publication to be held in Chicago in September 2021 (https://peerreviewcongress.org). The closing date for abstract submissions is January 2021.

Veterinary Medical Writing

SECTION EDITOR

Karim Montasser
dr.karimmontasser@gmail.com

Editorial

All three articles in this Veterinary Medical Writing section were written when COVID-19 was not classified as a pandemic. Yet all three articles tie into the current narrative of regulation in veterinary medicine and the need for communication between human and veterinary medicine. This could be seen as a coincidence, but I like to think that the members of EMWA are ahead of their time when it comes to trends in medical communication. I will let you be the judge. Thank you very much to our regular contributors Beatrix Doerr and Cemile Jakupoglu and a special thank you to Henry Smith who was kind enough to give us valuable insight into "One Health" as it is implemented in Japan.

Karim Montasser

Research in veterinary medicine takes off!

Do you remember my September 2019 article where I wrote about the need for data sharing in veterinary medicine and suggested that veterinary research could learn from the experience gained in humans?[1]

Meanwhile, veterinarians have taken this idea even further. It gave me great pleasure to learn that in November 2019, the Texas A&M University College of Veterinary Medicine & Biomedical Sciences and the University of Washington School of Medicine launched a large project studying ageing in dogs. This project goes beyond data sharing: it intends to create a community for dog owners, researchers, and volunteers. Dogs are nominated for participation by their owners, and data will be collected via questionnaires and the sharing of veterinary medical records. The project plans to include 10,000 dogs in an open-data platform that can be accessed by scientists all around the world. It aims to contribute to the knowledge about ageing in dogs and – with that – is expected to also shed light on ageing in humans.[2]

Another exciting project relying on the participation of pet owners is the Darwin's Ark project. It assesses the influence of genetics on health and behaviour. So far, more than 25,000 dogs have already been registered, and a similar project with cats is planned to be launched soon.[3]

Do you know of any similar veterinary projects involving data sharing, use of veterinary medical records, pet owner engagement, or other novel ways to gather data? Please do let me know!

References

1. Doerr B. Go, vets. Go! Med Writ. 2019;28:84–5.
2. Texas A&M, Washington Launch Historic Dog Aging Project. [cited 2020 Jan 13]. Available at: https://vetmed.tamu.edu/news/press-releases/texas-am-washington-launch-historic-dog-aging-project/
3. Join the world's largest pet citizen science project [cited 2019 Dec 03]. Available at https://darwinsark.org/

Acknowledgement

The author thanks Jessica Lin for her language editing.

Beatrix Doerr
beatrix.doerr@gmail.com

Regulatory in veterinary medicine takes off!

As announced in the December 2019 *Medical Writing* regulatory news section, the new veterinary regulation was published in January 2019. It repeals Directive 2001/82/EC, while the equivalent Directive 2001/83/EC for human pharmaceuticals is still in place. Thus, this is the first time the European Council adopted a medicines regulation addressing veterinary medicinal products in particular. As stated in the preamble, the new regulation aims to adapt "*the regulatory framework for veterinary medicinal products to scientific progress, the current market conditions, and economic reality, while continuing to ensure a high level of protection of animal health, animal welfare and environment and safeguarding public health*".[1]

The new regulation shall meet the specific needs of the veterinary sector, which differs substantially from the human sector in its smaller commercial potential yet having additional considerations of a diversity of animal species and animal therapeutic needs.

Antimicrobial resistance is a global public health concern, and, therefore, in accordance with the "One Health" approach, prudent use of antimicrobials is required. In October 2019 the European Medicines Agency (EMA) published three main criteria that were recommended for the selection of antimicrobials to be restricted to human use, namely high importance to human health, risk of resistance transfer, and low importance to animal health.[2]

Changes in the new directive that are generally welcomed from an industry perspective include the removal of the requirement for renewals of marketing authorisations (MA) that will now be valid for an unlimited period) and opening up of the centralised procedure to more kinds of products (it will be open to any application for which an MA has not previously been granted in the EU through the National, Mutual Recognition or Decentralised Procedure).[3] Other changes, such as the harmonisation of Summary of Product Characteristics, might result in a higher workload.

Advice on how to meet data requirements for novel therapies (such as gene therapy, regenerative medicine, tissue engineering, blood product therapy, phage therapy, nanotechnologies) was published by the EMA in August 2019,[4] so vets go ahead!

There are plenty of other changes not detailed in this short insight. Will the new veterinary legislation succeed in meeting the aforementioned aims of the European Council? Let's hope for the best!

References

1. Regulation (EU) 2019/6 of the European Parliament and of the Council of 11 December 2018 on veterinary medicinal products and repealing Directive 2001/82/EC. [cited 2020 Jan 31]. Available at: https://eur-lex.europa.eu/eli/reg/2019/6/oj
2. Advice on implementing measures under Article 37(4) of Regulation (EU) 2019/6 on veterinary medicinal products – Criteria for the designation of antimicrobials to be reserved for treatment of certain infections in humans. [cited 2020 Jan 31]. Available at: https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/advice-implementing-measures-under-article-374-regulation-eu-2019/6-veterinary-medicinal-products-criteria-designation-antimicrobials-be-reserved-treatment-certain_en.pdf
3. Bate, K. The new veterinary medicines regulation: What we know and what we don't. Animal Pharm, January 28, 2019. [cited 2020 Jan 31]. Available at: https://www.cyton.com/sites/default/files/New-EU-Veterinary-Medicines-Regulations_Cyton.pdf
4. Advice implementing measures under Article 146(2) of Regulation (EU) 2019/6 on veterinary medicinal products – Scientific recommendation on the revision of Annex II to Regulation (EU) 2019/6 on veterinary medicinal products. [cited 2020 Jan 31]. Available at: https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/advice-implementing-measures-under-article-1462-regulation-eu-2019/6-veterinary- medicinal-products-scientific-recommendation-revision-annex-ii-regulation-eu-2019/6-veterinary-medicinal-products_en.pdf

Acknowledgement

The author thanks Pascale Canning for language editing.

Cemile Jakupoglu
cjakupoglu@cyton.com

A tale of "One Health" from Japan:

Veterinarians in the fight against severe fever with thrombocytopenia syndrome

"One Health" – the concept of optimal health for animals, humans, and the environment – is a familiar term to veterinarians, but how well is it know to the wider medical community? I used to have a rather sketchy idea of this concept, but that has been rectified since I started teaching Medical English at a vet school in Japan. Reading up on One Health in Japan for my job, I have come across a fascinating story on the role of veterinarians here in combating one particular emerging infectious disease. This story concerns a zoonotic viral disease first reported in China that currently lacks a cure, but it has nothing to do with COVID-19. The disease in question is called severe fever with thrombocytopenia syndrome (SFTS), and efforts to counteract it have benefitted from a truly wide collaboration.

Figure 1. Deaths of two zoo-reared cheetahs due to SFTS announced on Japanese television news (All-Nippon Network News, August 18, 2017)
The caption on the bottom of the screen reads "At a Zoo: Ticks? Two cheetahs dead".

What is STFS?

SFTS is dwarfed by COVID-19 in terms of geographical reach and also has a different viral lineage. It is an arthropod-borne banyangvirus seen in regions of China, South Korea, and western Japan where certain tick species are active. Despite its low prevalence, SFTS causes concern due to its high fatality rates, reportedly ranging from around 20% in humans to 100% in cheetahs.[1]

The first case in Japan was reported in 2012, when a patient in her 50s was hospitalised for fever, fatigue, vomiting, and diarrhoea, with low WBC and platelet counts, and tragically succumbed one week later.[2,3] Her doctors suspected viral infection, but they turned to the laboratory of veterinary infectious disease specialist Dr Ken Maeda at Yamaguchi University for help in determining the cause of death. Dr Maeda's team succeeded in isolating the virus in Vero and feline embryo cells, and the pathological diagnosis of SFTS for this human patient was subsequently confirmed through genome sequencing and real-time PCR analysis at other veterinary specialist laboratories. The rapid isolation and identification of the virus and the classification of SFTS as an emerging infectious disease in Japan became a significant news story here.

Crossing species boundaries

Three years later, two animal patients sadly made headlines. They were cheetahs at Hiroshima's Aso Zoo that died in quick succession, with SFTS as the pathological diagnosis determined from tissue sample analysis (Figure 1). Tick bites were the probable cause of the first case, but cheetah-to-cheetah transmission could not be ruled out for the second case. However, the diagnosis of SFTS allowed zoo veterinarians to take appropriate action: the surviving cheetahs and other animals at the zoo were treated with acaricides and have remained free of SFTS ever since.

Transmission routes have also been questioned for human patients since some cases cannot be attributed to tick bites. Attention has

Figure 2. Public information pamphlet on SFTS for pet owners produced by the Kyoto Prefectural Government
This pamphlet educates pet owners on risks and prevention for this tick-borne disease.

thus been focused on the search for other vectors and sentinel species. For example, in one Japanese prefecture, SFTS seroprevalence in raccoons increased from 0% to nearly 50% between 2007 and 2016, foreshadowing the emergence of multiple human cases towards the end of that period. Similar surveys have been conducted for wild boar and deer, as well as stray cats and dogs.

The hunt for transmission routes has taken some surprising turns. One case report in 2019 involved a veterinarian – as the (fully recovering) patient. He had treated and necropsied three SFTS-positive cats and may have been infected through the exposure of the eye membrane to aerosol particles. Dogs show milder SFTS symptoms than cats, but they are seemingly another vector species. In 2017, national television news reported the transmission of SFTS from a pet dog to his owner through salivary contact with the eye membrane. Viewers were quickly reassured that "both the man and his pet have recovered".[5] Clearly a truly multidisciplinary medical team is required when two residents of the same house have the same disease but belong to different species.

A model for the One Health approach

The case of SFTS in Japan has been cited as a model of how a One Health approach should work,[3] and it is easy to see why. Evidence has come from a staggeringly wide range of sources: practitioners in human medicine, veterinarians specialising in small animals, zoo animals, and wildlife population surveys, experts in virus isolation and sequencing, infectious disease modelling, and tick biology, and public health officials (to name but a few). Following a comprehensive evaluation, this evidence has formed the basis for informational campaigns targeting both health professionals and the wider public. Armed with this information, physicians can now make quicker diagnoses of SFTS for both human and animal patients, and the public is better educated on how to avoid the risks to themselves and their pets (Figure 2). This has been achieved without creating irrational panic. Thanks also to the multidisciplinary efforts, progress is being made towards treatment: a novel agent (favipiravir) has shown efficacy in mice,[6] and early promise in humans.[7] What we have learned about SFTS in people, cats, dogs, raccoons, and cheetahs will ultimately benefit people, cats, dogs, raccoons, and cheetahs.

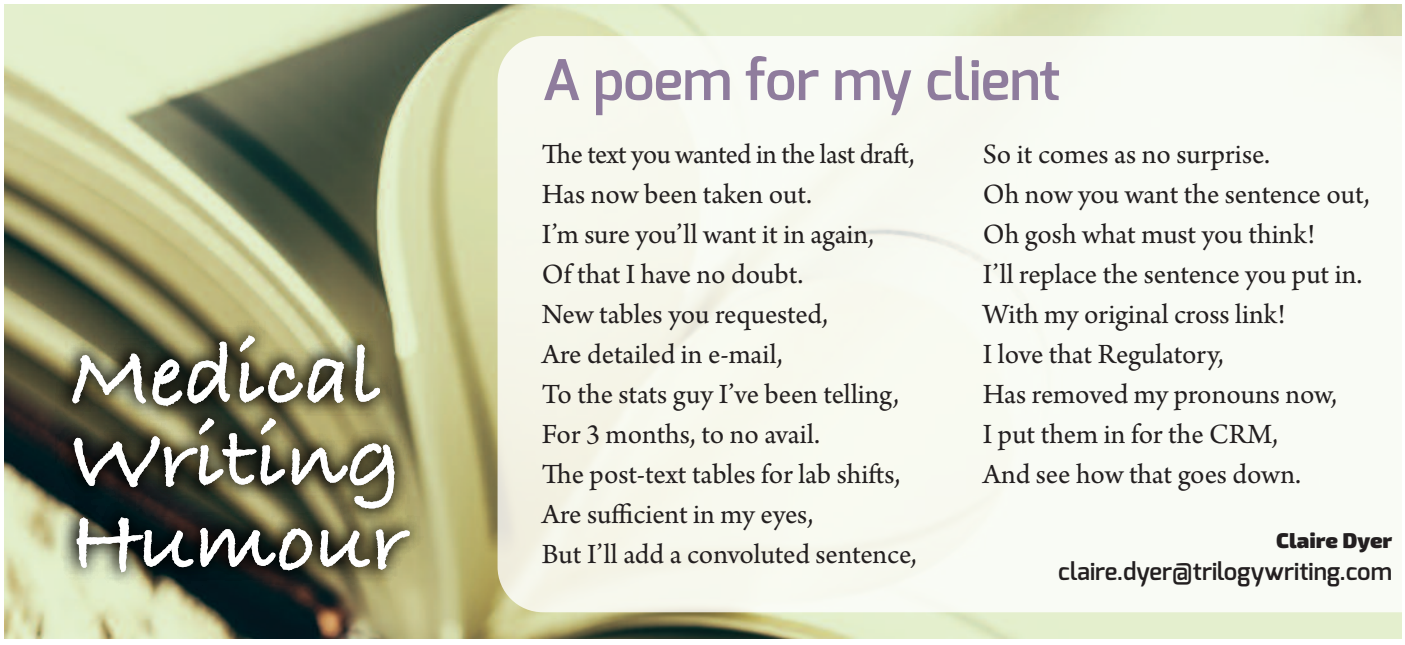
Broader implications of a One Health approach

Among the many morals of this story, I want to highlight one key message for us. One Health means that veterinary medical writing – and any other part of medical communication for that matter – belongs in the mainstream. We should be open to the idea that crucial scientific evidence can come from anywhere.

References

1. Maeda K. Recent findings for SFTS virus. In Proceedings of a Symposium on Collaboration for One Health, Oita, Japan, 2019. Japanese. Available from https://www.mhlw.go.jp/file/06-Seisakujouhou-10900000-Kenkoukyoku/0000196828.pdf
2. Shimoda H, Kuwata R, Maeda K. Severe fever with thrombocytopenia syndrome viewed from a veterinary perspective. Modern Media. 2016;62:23–30, 2016;68(1). Japanese
3. Takahashi T. Medical and Veterinary Collaboration against Severe Fever with Thrombocytopenia Syndrome. The Medical Frontline. 2017;72:579-83. Japanese
4. Kida K, Matsuoka Y, Shimoda T, et al. A case of cat-to-human transmission of severe fever with thrombocytopenia syndrome virus. Jpn J Infect Dis. 2019 Sep;72(5): 356–8.
5. Dog-to-owner transmission of a tick-borne infection confirmed in Japan. All-Nippon News Network. 2017 Oct 11.
6. Tani H, Komeno T, Fukuma A, et al. Therapeutic effects of favipiravir against severe fever with thrombocytopenia syndrome virus infection in a lethal mouse model: Dose-efficacy studies upon oral administration. PLoS ONE. 2018 Oct;13(10):e0206416.
7. Song R, Chen Z, Li W. Severe fever with thrombocytopenia syndrome (SFTS) treated with a novel antiviral medication, favipiravir (T-705). Infection. 2019; https://doi.org/10.1007/s15010-019-01364-9

Henry Smith
Joint Faculty of Veterinary Medicine
Kagoshima University
Kagoshima, Japan
K5908476@kadai.jp



A poem for my client

The text you wanted in the last draft,
Has now been taken out.
I'm sure you'll want it in again,
Of that I have no doubt.
New tables you requested,
Are detailed in e-mail,
To the stats guy I've been telling,
For 3 months, to no avail.
The post-text tables for lab shifts,
Are sufficient in my eyes,
But I'll add a convoluted sentence,
So it comes as no surprise.
Oh now you want the sentence out,
Oh gosh what must you think!
I'll replace the sentence you put in.
With my original cross link!
I love that Regulatory,
Has removed my pronouns now,
I put them in for the CRM,
And see how that goes down.

Claire Dyer
claire.dyer@trilogywriting.com

Good Writing Practice

Non-contiguity: Adjective clause

Introduction

An adjective clause displaced from its modifee by an intervening syntactic unit is a distraction. Another distraction is the vague adjective clause that seems to refer to an entire sentence rather than to a definite modifee. Such vagueness has resulted in the pejorative term the "vague which".

A modifee is a syntactic unit modified by a modifier. It is not cited by even unabridged dictionaries but should be because of its succinctness. Further justification for its usage is by analogy to conventional pairs: *employer.employee; mentor.mentee*.

Experimental sections

Part 1 – Results section: Result (observation)

Example: adjective clause non-contiguity

The appliance resulted in the characteristic tooth movement for control groups, which consisted of three phases.

Revision 1

***For control groups**, the appliance resulted in the characteristic tooth movement, which consisted of three phases.*

Revision 2

*For control groups, the appliance resulted in the characteristic **three-phase** tooth movement.*

Notes

The contiguity of *groups* and the plausibility of *which consisted of* as its modifier renders the distraction an impeded immediate comprehension. In Revision 1, not only does the transposition of *for control groups* enable contiguity of the adjective clause to its modifee, but it also enables the transposed prepositional phrase *for control groups* to function as a sentence orientation.

In Revision 2, the adjective clause is syntactically reduced into the attributive compound adjective *three-phased*, the succinctness of which renders Revision 2 as a useful option. However, the complete adjective clause because of its length and sentence end-position placement is more emphatic than the compound adjective *three-phrase*.

Part 2 – Results section: Result (sequential observations)

Example: adjective clause non-contiguity

Fluid was displaced into the joint cavity, which increased regional synovial fluid pressure.

Revision 1

*Fluid was displaced into the joint cavity, a **displacement** that increased regional synovial fluid pressure.*

Revision 2

*Fluid was displaced into the joint cavity, **increasing** regional synovial fluid pressure.*

Revision 3

*Fluid was displaced into the joint cavity, and **regional synovial fluid pressure was increased**.*

Revision 4

*The **displacement** of fluid into the joint cavity **increased** regional synovial fluid pressure.*

Notes

In the Example, the vague *which* seems to modify the entire preceding independent clause rather than one specific noun. The adjective clause could modify *fluid*, but only the fluid that was displaced into the joint cavity, not just fluid.

In Revision 1, the longest of the three revisions, the intended modifee is emphasised by the usage of a noun derivative (*displacement*) of the verb *displaced*. Although *displacement* is an explicit modifee of the adjective clause, the revision seems overly obvious (a hypercorrection), redundant, and usually not preferred to the Example.

In Revision 2, the participle *increasing* modifies the whole sentence as does the adjective clause in the Example, but without the backtracking of the relative pronoun *which*. One distraction of *increasing* is a misagreement in tense to a past observation, but its succinctness and fluidity outweigh its disadvantages and is consistently preferred to all the other revision options.

In Revision 3, the compound sentence befits the two observations but lacks the fluidity of Revision 2.

In Revision 4, the thematically focused after-the-fact subject *displacement* depends on a prior mention in the text and therefore more appropriate for the Discussion section; however, it does contain *the* appropriate past tense.

SECTION EDITORS

Wendy Kingdom
info@wendykingdom.com

Amy Whereat
amy.whereat@speakthespeech.fr

Contextual sections

Part 1 – Introduction section: Research problem – pertinent background

Example: adjective clause non-contiguity

There are several methods to estimate the CIR that are more efficient than those previously used.

Revision 1

***To estimate the CIR**, there are several methods that are more efficient than those previously used.*

Revision 2

*There are several **CIR-estimating** methods **that** are more efficient than those previously used.*

Notes

It is not clear whether the adjective clause *that are more efficient than those previously used* modifies *CIR* or *methods*. Although the plurality of *that are* relates to the adjective clause to the modifee *methods* and not to *CIR*, a reader may be uncertain whether the author committed a grammatical mistake of subject-verb misagreement in number.

Revision 1 involves transposing the displacing unit (infinitive phrase: *to estimate the CIR*) to the sentence-orientating position enabling the adjective clause to be contiguous to its modifee *methods*.

Revision 2 involves syntactically reducing the intervening infinitive phrase into a compound-noun pre-modifier *CIR-estimating*, enabling contiguity of the adjective clause to its modifee. The usage of this revision may depend on the prior mention of *CIR-estimating methods* because

it is an attenuated form of *methods to estimate CIR*.

Overall, the infinitive phrase displacement distraction (and its revisions) – similar to that for the prepositional phrase-caused displacement in Example 1 – is an example of an adjective clause modifying a specific modifee.

Part 2 – Introduction section: Research problem – pertinent background

Example: adjective clause modifying a sentence

Some Alcyonarian soft corals release toxins into seawater, which affect community composition and function.

Revision 1

Some Alcyonarian soft corals release ***into seawater*** *toxins, which affect community composition and function.*

Revision 2

Some Alcyonarian soft corals release toxins into seawater, affecting community composition and function.

Revision 3

Some Alcyonarian soft corals release toxins into sea water and thereby affect community composition and function.

Revision 4

The ***release*** *of toxins into sea water by some Alcyonarian soft corals affects community composition and function.*

Notes

What is the modifee of the *which* clause? As indicated by the plural verb *effect*, the modifee can be the noun phrase *toxins* or *toxins into seawater* or the whole independent clause. This usage of the adjective clause is similar to that in Example 1 but *toxins* is a likely modifee for the adjective clause, so the somewhat awkward transposition of *into seawater* enables contiguity of the adjective clause (Revision 1).

In Revision 2, syntactic reduction of the adjective clause to the participial phrase *affecting* may be a preferred option, because of its smooth flow, appropriate present tense for known information, and succinctness. Similar to the adjective clause, the participial phrase is probably modifying the displaced noun *toxins*.

In Revision 3, coordinating of independent clauses, the longest revision, explicitly coheres the two relations. In Revision 4, *release* as the subject of the sentence is stated after the fact, that is, occurring in a section of a journal article (the Discussion) after *Some Alcyonarian soft corals release toxins into seawater* was stated in a prior section (the Results).

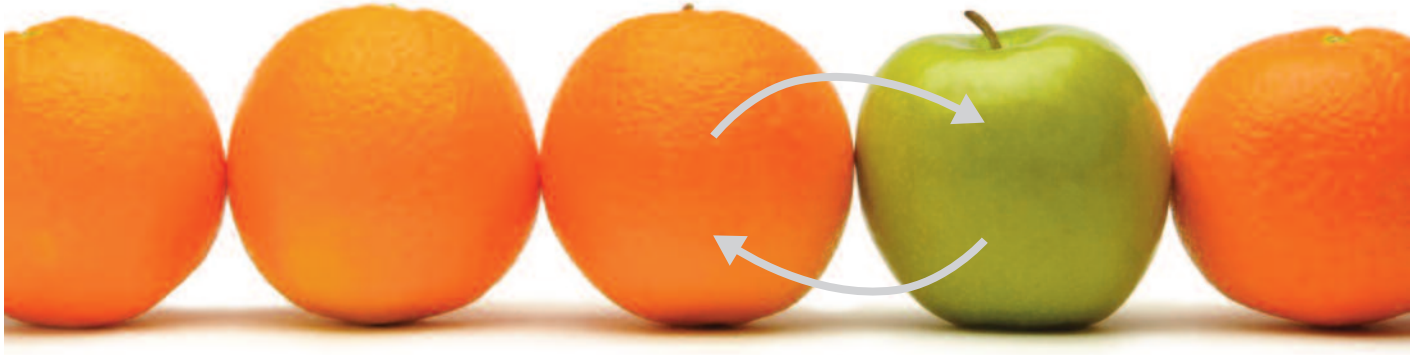
Summary

The four examples are equally distributed between Experimental (Results) and Contextual sections (Introduction) of a journal article, indicating a lack of section specificity.

A modifier displaced from its modifee is distracting by impeding immediate comprehension. In contrast, a modifier of a whole sentence because of its conventionality is just a dissonance. For an adjective clause modifying a specific modifee, revision involves eliminating the distance between modifier and modifee by transposing the intervening syntactic unit to the sentence-initial position or transposing a pre-modifier form of the modifier so that modifier and modifee are contiguous.

For an adjective clause modifying an entire sentence (the vague *which*), the adjective clause can be reduced into a participial phrase or expanded into a coordinate independent clause. Each of the revisions expresses a different nuance.

Michael Lewis Schneir, PhD
Professor, Biomedical Sciences, Ostrow School of Dentistry of The University of Southern California, Los Angeles, CA
schneir@usc.edu



Schematised distractions and preferred revisions

Transposition of a disruptive prepositional phrase

The appliance resulted in the characteristic tooth movement for control groups, which consisted of three phases.

→ ***For control groups****, the appliance resulted in the characteristic tooth movement, which consisted of three phases.*

Transposition of a pre-modifier form of an adjective clause

There are several methods to estimate the CIR that are more efficient than those previously used.

→ *There are several* ***CIR-estimating*** *methods that are more efficient than those previously used.*

Syntactic reduction to a participial phrase

Fluid was displaced into the joint cavity, which increased regional synovial fluid pressure.

→ *Fluid was displaced into the joint cavity,* ***increasing*** *regional synovial fluid pressure.*

Some Alcyonarian soft corals release toxins into sea water, which effect community composition and function.

→ *Some Alcyonarian soft corals release toxins into seawater,* ***affecting*** *community composition and function.*

EMWA 2020
Virtual Autumn Conference

The Autumn conference is not cancelled. Instead we will be holding a virtual conference.

Due to the ongoing COVID-19 pandemic, EMWA's Executive Committee has decided to shift the Autumn conference this year to a virtual format.

The virtual Autumn conference will be held in November, although exact dates have not yet been selected. EMWA's Executive Committee, Professional Development Committee, and Head Office are currently working to deliver a live and interactive conference experience that you can attend from the safety of your own home or office.

EUROPEAN MEDICAL WRITERS ASSOCIATION

The virtual Autumn conference will feature the usual conference events, including:

- Workshops
- Symposium
- Opening session
- Freelance Business Forum

Out on Our Own

SECTION EDITOR

Laura A. Kehoe
laura.a.kehoe@gmail.com

Welcome readers,

As I write this editorial, the coronavirus pandemic is having profound effects around the globe. In a time of uncertainty, stress, and fear, I try to think of the positives. For me, one positive is that I'm a freelancer. I have not lost any clients, I can still work, I have my desk already set up at home, and I can work to a flexible schedule; wonderful as I now have two young children at home full time! We are all going through this pandemic together, but everyone's situation is different. Actually, positively thinking, freelancers could be at an advantage during this time. Many full-time workers, who usually work in the workplace, are in unknown territory by setting up their offices at home. Perhaps this is easier if you don't have children that you need to homeschool, occupy, and comfort, but for people in this situation, juggling full-time work with children and a partner's full-time work is near impossible. Many companies will see a decline in work productivity, but that's where us freelancers can "fill in the gaps". We're already set up to do exactly what many companies are asking their employees to do. Stay open-minded, network virtually to adhere to the social distance rules, and who knows what work may come your way during this time!

Our author this issue, Archana Nagarajan only decided to become a freelancer a year ago, and by the sounds of it, it was the best decision for her. She explains about her quick journey into becoming a freelance medical writer and the projects she's taking on during that time. Hungry to learn all fields of medical writing, she expanded her medical communications repertoire to include regulatory writing as well. As we've sadly missed the May congress, and thus the Freelance Business Forum, she remains passionate about sharing knowledge and does so by offering some essential tips for new freelancers starting out. Maybe she'll expand on these at the next EMWA!

Stay safe until then.

Laura A. Kehoe

Switching careers: My path to medical writing and freelancing

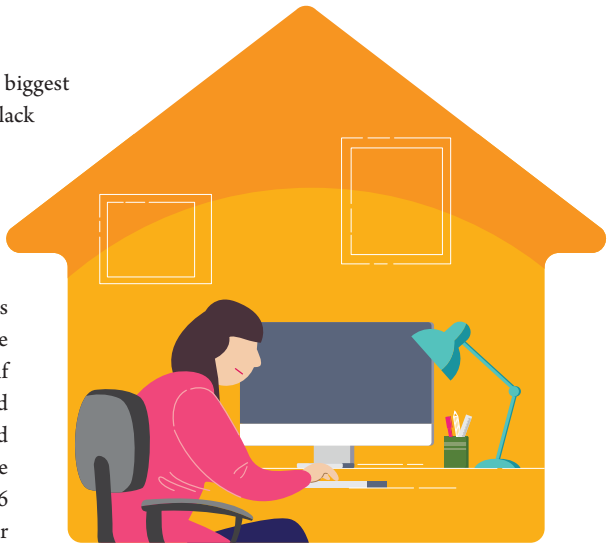
I get to share my journey into freelance medical writing at a time when the world has come to a standstill. We all are learning new ways of balancing our professional and personal lives.

It has been a year since I became a freelance medical writer and the journey, in some ways, reflects and resonates with so many of my fellow medical writers, but it is also my own unique, personal journey. I currently work in medical communications as well as regulatory writing: editing and writing journal articles for non-native speakers and also, writing regulatory documents for medical devices.

Medical writing as a career

I did my PhD in genetics from Bengaluru, India, and moved to Nice, France, for a post-doctoral stint in 2009. Ten years, four countries (France, England, Norway, and Germany), three post-doctoral positions in ageing biology and one child later, I am now settled in the Hanseatic city of Hamburg in Northern Germany.

After taking time off for a couple of years to raise my daughter and helping with family sickness, I wanted to get back to work. I absolutely love science and immensely enjoyed my career as a scientist. However, I had reached a point where doing more post-doctoral research did not make sense. The biggest hurdle at this point was the lack of knowledge and guidance for an alternative career. I spent a year in limbo unable to make up my mind on what to do. Germany is not an easy country to be employed in, especially if your degrees are not earned here, you have never worked here or if you do not speak the language (I did spend 6 months during this 1 year upgrading my German language skills). I had always enjoyed the process of science communication and was looking at ways of leveraging that. Around this time, I shared my predicament with one of my close friends, and as it turned out, she was working as a medical writer in the USA! I think that life is full of such moments when a path becomes clearer at just the right time. She shared her journey into medical writing with two parental breaks to raise her kids. Following her suggestion, I joined EMWA in December 2018, and since then, there has been no looking back.



EMWA has plenty of resources online and also in the quarterly *Medical Writing* journal. I read through the articles available on various types of medical writing (the primer on *Introduction to Medical Writing* available on the website is pure abbreviated gold). I did not want to move out of Hamburg, where my family is settled, and where we have friends now. Hamburg does not have too many pharmaceutical firms and the medical writing jobs are limited. Freelancing, thus, seemed to be the logical way forward.

Medical communications: A logical extension from being a scientist

I started with writing and editing manuscripts for journals as this is a skill, we all learn as scientists. However, I quickly realised that writing for someone else was not the same! Most importantly, you are not intricately aware of the science behind every experiment. Second, journal selection, sticking to various guidelines (ICMJE, ISMPP, CONSORT, etc.) are some of the non-scientific work that also takes quite a bit of time. However, having a PhD does have its advantages. I have, over time, become adept in picking up new knowledge with ease, doing literature searches, extracting information from the thousands of scientific papers available online, and judging a paper for its merits and shortcomings. Over the last one year of freelancing, I have written and edited manuscripts in life sciences (both preclinical and clinical) and I should say that learning about new therapeutic areas has been very rewarding.

Medical communication was a natural choice as it was an extension of my knowledge as a scientist and involved dissemination of science to a varied audience. In the past year, I have also written quite a few articles and blogs for various online magazines and those have been extremely wonderful experiences (a few links are at the end of the article). I wrote for a Women in Science series on four Indian women scientists, which was a moment of pride for me to bring to the world the stories of these wonderful superwomen from my country.

EMWA conference and networking: A brilliant experience

I attended the EMWA spring conference in Vienna last year, which was an invaluable experience and also, it reaffirmed my belief in choosing medical writing as a career. I benefitted immensely from all the workshops that I attended, the seminar on medical writing as a career, the freelance business forum and most importantly, the network of medical writers that I could learn from. The biggest gain for me was to meet other medical writers who were starting out like me and taking those tentative steps towards a new career. It helped me get over my imposter syndrome. I eventually attended the Malmö conference in November 2019 as well and obtained the foundation certificate issued by EPDP.

Regulatory writing: A different world

Vienna is where I also networked with the owner and co-founder of the company that is currently one of my clients. I had mentioned (probably more than once!) to him that I really wanted to get into regulatory writing and thus, was looking for an opportunity. He reached out to me sometime in August to work as a consultant for FDA regulatory submissions for medical devices and combination products. I was really happy to join the team, and in the past 6–7 months, my learning curve has been on a steep upward trend – learning about various FDA regulatory requirements, writing QMS documents, interacting with clients, working remotely in a team (that are worldwide and in all the possible time zones!), learning the differences in FDA requirements and EU-MDR, project management – the list is endless. I have heard many times over that regulatory writing is dry. Till now, my experience has been the opposite. I find it challenging and interesting. Coming from a basic sciences background, I find regulatory writing fascinating. I'm learning about mechanisms that control the safety and efficacy of the product, how the life cycle of an entire drug or device is controlled while making sure that the latest research is translated into medical innovations that benefit the end users without too much of a delay, and how clinical trials are conducted.

I enjoy the flexibility that my freelancing offers, yet I do find that I need to be on my toes and be ready to adapt to client needs and also, keep updating my knowledge base. However, I have no complaints during these unprecedented times that we all find ourselves in. I already work from home, and focus-wise it is not much of a shift. I feel that I made the right decision to be my own boss.

A few of the articles that I wrote last year:

- https://medium.com/sci-illustrate-stories/janaki-ammal-466b644a4369
- http://www.sciwri.club/archives/9162
- https://medium.com/sci-illustrate-stories/bibha-chowdhuri-c7c48792d2b1
- https://medium.com/sci-illustrate-stories/asima-chatterjee-1ca581dc542f

Freelancing as a medical writer and consultant

I thought I would share some tips and advice for starting and sustaining as a freelance medical writer.

- **Network, network and network.** Many of us who come from a scientific background are introverts and clearly feel like a fish out of water when it comes to networking. However, there is no other way to sustain your business than to keep networking (especially in the beginning). You have to put yourself out there!
- **Update your LinkedIn profile.** Don't have a generic one. Have one that stands out. List the services that you offer. Actively engage in the platform (network!).
- **Don't be overwhelmed.** We all were beginners once. Don't expect miracles from day one.
- **Update your knowledge constantly.** I have signed up for some newsletters that keep me posted about the happenings in the pharma world, medical devices, etc.
- **Join EMWA and their local city groups.** Yes, it is worth it! Feeds back to my point number one.
- **Take up a few courses.** Even if it means spending some money (if you can). It will be worth it in the end. I did some courses on scientific writing from Coursera, Edx, etc.
- **Set up a nice workspace.** Keep it organised and clutter-free.
- **Get to know the country-specific laws.** Learn about freelancing laws in the country you reside in. Always have your tax papers, invoices, etc., in order to avoid a last-minute scramble.
- **Remember you are in it for the long run.** Keep calm and enjoy writing! Enjoy the luxury of going out on a walk, able to do a workout, cook with the music on in between writing when you work from home!

Archana Nagarajan
Freelance Medical Writer and Regulatory Consultant
Hamburg, Germany
archananag0611@gmail.com

Upcoming issues of **Medical Writing**

September 2020:

European Union regulations

This issue will focus on new EU regulations and their impact on medical writing. Key topics will include changes to centralised procedures, effects of Brexit on the EMA, and new regulations on medical devices, drug-device combinations, and veterinary medicines.

Guest Editor: Ana Madani

This issue is closed to new contributions.

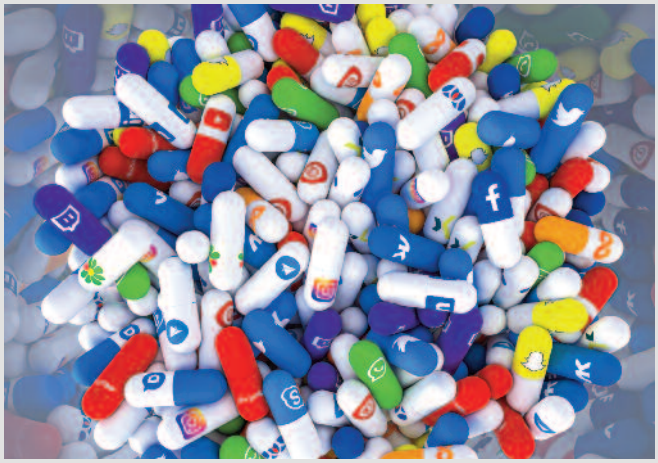
December 2020:

Writing for patients

This issue will feature articles from some of the key opinion leaders in the area of writing for patients. We will cover aspects such as the current state of information given to patients and how we can do this better, the role of the medical writer with patient associations, the patient voice in research publications and writing up patient-reported outcomes, writing for the internet, and how patient needs are being incorporated into traditional medical communications.

Guest Editor: Lisa Chamberlain James and Amy Whereat

The deadline for feature articles is September 8, 2020.



March 2021:

Social media

For many people social media has become a primary source of information, including that related to medicine and healthcare. This issue will include articles about this trend, how to leverage the different social media tools, and how to write for social media.

Guest Editor: Diana Ribeiro

The deadline for feature articles is December 8, 2020.

CONTACT US

If you have ideas for themes or would like to discuss any other issues, please write to **mew@emwa.org**.

http://journal.emwa.org/