The CODE-EHR global framework – lifting the veil on health record data

Folkert W. Asselbergs[1,2], Dipak Kotecha[3,4]

1) Amsterdam University Medical Center, Department of Cardiology, University of Amsterdam, Amsterdam, The Netherlands

2) Health Data Research UK and Institute of Health Informatics, University College London, London, UK

3) Institute of Cardiovascular Sciences, University of Birmingham, Birmingham, UK

4) Health Data Research UK Midlands, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK

Acknowledgement:

Folkert Asselbergs is supported by UCL Hospitals National Institute for Health and Social Care Research (NIHR) Biomedical Research Centre. Dipak Kotecha is supported by the NIHR Health Technology Assessment programme (NIHR130280) and the NIHR Birmingham Biomedical Research Centre (NIHR203326). Both Folkert Asselbergs and Dipak Kotecha received support from the EU/EFPIA Innovative Medicines Initiative BigData@Heart related to this described work (grant no. 116074). The opinions expressed in this paper are those of the authors and do not represent any of the listed organizations.

Word count: 882; References x 10; Figures x 1.

With the emergence of electronic health record (EHR) systems, the application of big data research, including support from artificial intelligence, has the potential to massively enhance the value of routine data in healthcare. Big data research can also play a crucial role in generating evidence for clinical practice guidelines. One major advantage is the dynamic provision of contemporary evidence, which can update policies often based on historical studies that do not reflect current disease burden or modern management strategies. Additionally, big data analysis can address patients with multimorbidity that are not well represented in clinical trials that underpin practice guidelines. However, there are several major obstacles and limitations to using routine care data, such as data access and quality, privacy and security. Data are often not standardized, with limited interoperability between data sources. The COVID-19 pandemic has illustrated the need for open-access contemporary healthcare data, and the urgency to develop new approaches to improve data anonymization and pseudonymization whilst preserving transparency and traceability of data. The retractions of COVID-19 studies using the Surgisphere registry [1,2] are a clear signpost that guidance is needed to build trust using EHR data for evidence generation.

Transparency of data handling is of critical importance to report trustworthy results obtained from routine care data and enable validation across different clinical settings. Clarity from a broad range of stakeholder perspectives is critical to direct this new scientific discipline, including regulators, healthcare professionals, journal editors, payers and, most importantly, the general public. Lack of public trust is detrimental to big data research and the willingness to share healthcare data, and adherence to data privacy laws does not guarantee societal approval. Studies indicate that individuals and the general public are cautiously supportive of sharing their health information, but have worries about data misuse and commercial

exploitation [3]. Obtaining a social license can boost big data initiatives by fostering accountability and sustainability in health data sharing while still adhering to existing regulations. Engaging early with patients and the public in the research process can establish a solid foundation for big data research studies and produce more relevant results to society.

It is important to emphasize that EHR systems were originally not developed for research purposes (with corresponding strict clinical definitions and standards) but built for business intelligence usage. Heterogeneity in phenotype definitions therefore exists between EHR systems and also between regions, countries and disease domains which inherently leads to variation in data quality that will affect any subsequent analysis. Another source of variation are missing data and the use of unstructured data such as free text within EHR systems. Variation in data collections has motivated national audit registries to start working on minimal datasets with only discrete variables and adopt standards to facilitate comparative research for quality of care. Similar limitations apply to research and therefore the ESC has embraced EuroHeart to foster collaboration across Europe to harmonize disease and outcome definitions [4]. Harmonization of phenotypes and outcomes are not only relevant to observational research - EHR data are now being used in clinical trials to support drug and device development and application, providing access to large patient numbers, unbiased follow-up, and near-indefinite monitoring of safety and durability of effect [5,6]. However, long-term follow-up may still be limited in some healthcare regions due to interoperability and governance issues that allow for linkage between hospital-based EHR systems, national outcome registries and/or primary care data.

All these challenges need to be known to readers of published studies, but also to researchers at the start of their study to plan ahead on solutions that can improve transparency and quality. Despite the availability of numerous reporting standards, there is no consensus on how to

realize the Findable, Accessible, Interoperable, and Reusable (FAIR) principles in the context of structured routine healthcare data [7]. A quality framework is needed to enhance the design and deployment of clinical research that depends on these increasingly crucial new sources of data.

To meet these challenges, an international stakeholder meeting was organized by the ESC and the BigData@Heart Consortium, including regulators (European Medicines Agency, US Food and Drug Administration), governmental agencies (European Commission, the UK National Institute for Health and Care Excellence, Innovative Medicines Initiative), leading medical journals (BMJ, European Heart Journal, The Lancet, Lancet Digital Health), and patient advocacy (European Heart Network, ESC Patient Forum), along with representatives from the pharmaceutical industry, payers and academic institutions. Following an iterative process and multiple workshops, the CODE-EHR best practice framework for the use of structured electronic health data was developed (**Figure 1**) to guide researchers and provide transparency on data construction, coding systems, analytical methods, information governance, and patient & public involvement [8,9,10]. CODE-EHR is more than a reporting checklist for publication purposes – it aims to guide investigators that are using routine healthcare data to design and conduct their study appropriately to achieve an optimal output. More importantly, use of CODE-EHR provides the opportunity to validate the results of studies by other researchers, achieving better implementation into the clinical care setting and the prospect of improving patient care. Adoption of this framework has now been included in the instructions to authors at the European Heart Journal and other major journals, helping to build trust among researchers, clinicians, regulators and eventually the public to fully embrace the opportunities provided by routine healthcare data.

References

1. MR Mehra, SS Desai, S Kuy, TD Henry, AN Patel. Retraction: Cardiovascular Disease, Drug Therapy, and Mortality in Covid-19. N Engl J Med. 2020;382:2582.
2. The Editors of the Lancet Group. Learning from a retraction. The Lancet. 2020;396:1056.
3. Aitken M, Cunningham-Burley S, Pagliari C. Moving from trust to trustworthiness: Experiences of public engagement in the Scottish Health Informatics Programme. Sci Public Policy. 2016;43:713-723
4. Wallentin L, Gale CP, Maggioni A, Bardinet I, Casadei B. EuroHeart: European Unified Registries On Heart Care Evaluation and Randomized Trials. Eur Heart J. 2019;40:2745-2749.
5. Kotecha D, DeVore AD, Asselbergs FW. Fit for the future: empowering clinical trials with digital technology. Eur Heart J. 2023;44:64-67
6. Wang X, Mobley AR, Tica O, Okoth K, Ghosh RE, Myles P, Williams T, Haynes S, Nirantharakumar K, Shukla D, Kotecha D. Systematic approach to outcome assessment from coded electronic healthcare records in the DaRe2THINK NHS-embedded randomized trial. Eur Heart J - Dig Health. 2022;3:426-436
7. Wise J, de Barron AG, Splendiani A, Balali-Mood B, Vasant D, Little E, Mellino G, Harrow I, Smith I, Taubert J, van Bochove K, Romacker M, Walgemoed P, Jimenez RC, Winnenburg R, Plasterer T, Gupta V, Hedley V. Implementation and relevance of FAIR data principles in biopharmaceutical R&D. Drug Discov Today. 2019;24:933-938
8. Kotecha D, Asselbergs FW, Achenbach S, Anker SD, Atar D, Baigent C, Banerjee A, Beger B, Brobert G, Casadei B, Ceccarelli C, Cowie MR, Crea F, Cronin M, Denaxas S, Derix A, Fitzsimons D, Fredriksson M, Gale CP, Gkoutos GV, Goettsch W, Hemingway H, Ingvar M, Jonas A, Kazmierski R, Løgstrup S, Thomas Lumbers R, Lüscher TF,

McGreavy P, Piña IL, Roessig L, Steinbeisser C, Sundgren M, Tyl B, van Thiel G, van Bochove K, Vardas PE, Villanueva T, Vrana M, Weber W, Weidinger F, Windecker S, Wood A, Grobbee DE; Innovative Medicines Initiative BigData@Heart Consortium, European Society of Cardiology, CODE-EHR international consensus group. CODE-EHR best practice framework for the use of structured electronic healthcare records in clinical research. Eur Heart J. 2022 Oct 7;43(37):3578-3588

9. Kotecha D, Asselbergs FW, Achenbach S, Anker SD, Atar D, Baigent C, Banerjee A, Beger B, Brobert G, Casadei B, Ceccarelli C, Cowie MR, Crea F, Cronin M, Denaxas S, Derix A, Fitzsimons D, Fredriksson M, Gale CP, Gkoutos GV, Goettsch W, Hemingway H, Ingvar M, Jonas A, Kazmierski R, Løgstrup S, Lumbers RT, Lüscher TF, McGreavy P, Piña IL, Roessig L, Steinbeisser C, Sundgren M, Tyl B, van Thiel G, van Bochove K, Vardas PE, Villanueva T, Vrana M, Weber W, Weidinger F, Windecker S, Wood A, Grobbee DE. CODE-EHR best practice framework for the use of structured electronic healthcare records in clinical research. BMJ. 2022;378:e069048

10. Kotecha D, Asselbergs FW, Achenbach S, Anker SD, Atar D, Baigent C, Banerjee A, Beger B, Brobert G, Casadei B, Ceccarelli C, Cowie MR, Crea F, Cronin M, Denaxas S, Derix A, Fitzsimons D, Fredriksson M, Gale CP, Gkoutos GV, Goettsch W, Hemingway H, Ingvar M, Jonas A, Kazmierski R, Løgstrup S, Lumbers RT, Lüscher TF, McGreavy P, Piña IL, Roessig L, Steinbeisser C, Sundgren M, Tyl B, Thiel GV, Bochove KV, Vardas PE, Villanueva T, Vrana M, Weber W, Weidinger F, Windecker S, Wood A, Grobbee DE. CODE-EHR best-practice framework for the use of structured electronic health-care records in clinical research. Lancet Digit Health. 2022;4:e757-e764

Figure Legends

Figure 1: The CODE-EHR framework. A minimum and preferred standards checklist for the use of structured electronic healthcare records in clinical research is available at https://www.escardio.org/bigdata and https://www.equator-network.org/reporting-guidelines/code-ehr-best-practice-framework-for-the-use-of-structured-electronic-health-care-records-in-clinical-research/.

CODE-EHR

Best practice framework for the use of structured electronic healthcare records in clinical research

BigData@Heart

ESC

www.escardio.org / bigdata

1. Dataset and linkage	Source of data Missing data approaches Completeness of follow-up How linkage was performed and its quality
2. Data fit for purpose	Origin and purpose of data Specify coding systems and classifications Quality assessment for data capture Outline all biases
3. Disease/ outcome definitions	Coding manual published prior to statistical analysis Detail on phenotyping algorithms Specify validation processes employed for coding schemes
4. Analysis	Details on statistical methods Links to any machine code or algorithms Internal validation of models External validation approaches
5. Ethics and governance	Patient consent processes Data privacy protection Patient and public involvement Process for obtaining data/code for verification/re-use