

Big Data Analytics in Adult Congenital Heart Disease: Why Coding Matters.

Folkert W. Asselbergs^{1,2,3}, Folkert J. Meijboom¹

Affiliations:

1. Department of Cardiology, Division Heart & Lungs, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands;
2. Institute of Cardiovascular Science, Faculty of Population Health Sciences, University College London, London, United Kingdom;
3. Health Data Research UK and Institute of Health Informatics, University College London, London, United Kingdom

With their study “Machine Learning Algorithms Estimating Prognosis and Guiding Therapy in Adult Congenital Heart Disease: Data from a Single Tertiary Centre including 10,019 patients”, Diller *et al.*¹ demonstrate that machine learning algorithms can tackle challenges within the complex field of congenital heart disease (CHD). They compared the outcomes of DL-algorithms in diagnosing, evaluating disease complexity and NYHA class to that of a test sample, and achieved an accuracy of 91.1%, 97.0% and 90.6%, respectively. These results are promising and the authors can be complimented with this “first in ACHD” use of these tools. The authors conclude that their study illustrates the applicability of machine learning algorithms in estimating prognosis and potentially even in guiding therapeutic management in adults with CHD (ACHD). Moreover, they state that these DL-algorithms can easily be scaled to multi-institutional datasets to further improve their accuracy and thereby ultimately may be used as online-based decision-making tools. These conclusions deserve some reflections.

We agree with the authors that there is a desperate need for data in ACHD, best illustrated by the ESC guidelines² that, apart from two exceptions, only contain level C recommendations. There are likely several reasons for this scarcity: (1) a lack of consensus regarding ACHD nomenclature, (2) absence of clinical care pathway harmonization across expert centers, (3) the heterogeneity of ACHD in terms of the initial diagnosis and (4) the diagnosis-, and sometimes institution- specific interventions that all have changed substantially over the years. Furthermore, complications of ACHD including heart failure, arrhythmia's and death often occur late in life, making it difficult to link outcome to the initial diagnosis and management.

Taken together, ACHD is not a rare disease, with an estimated 2.3 million and ± 1 million patients affected in Europe³ and the US⁴ respectively. In Europe, the management of these patients is currently scattered across many small, regional

hospitals, with only a few larger centers. However, even in the latter, patients with a specific diagnosis are too limited to apply sensible statistics. To overcome these issues, multiple national – the Dutch CONCOR registry (<http://concor.net>), the German Kompetenz Network (www.kompetenznetz-ahf.de) and Swedish Registry of Congenital Heart Disease (www.ucr.uu.se/swedcon) - and international registries including the Congenital Heart Surgeons Association (<https://www.echsacongenitaldb.org>) have been generated. Nonetheless, the “conventional way” of setting up and filling databases is labor-intensive and thereby expensive. Introducing electronic health records provides a unique opportunity to overcome this limitation by automatic data collection and – extraction from routine clinical care on a large scale. Future development of algorithms with the capacity of scanning digital patient files, structuring unstructured data (e.g. free text) and extracting symptoms, diagnoses, applied interventions, hospital readmission rates and long-term clinical outcomes, will lead to a considerable reduction in the costs of setting up- and keeping these registries up to date. Additionally, other data sources, currently lacking feasibility, including medical imaging-, sensor- and ECG data, could be implemented in these algorithms in the future. Having access and being able to analyze an amount of data that is far greater than used in conventional medical analyses, “Big Data”, from all imaginable sources that include both possible determinants for outcome as well as the outcome data itself, one might get a better grip on this extremely complicated and diverse patient category.

Figure 1: Possible sources of data for congenital heart disease using AI

There still are substantial practical issues that make generalization of this approach challenging in ACHD, as well as other diseases. Firstly, two different systems of coding ACHD diagnoses⁵ are still used in clinical practice, despite the ongoing effort of the International Pediatric Cardiology Coding Committee (IPCCC) for over 15 years. If one considers scaling DL to multiple institutions, consensus on diagnostic coding is essential to guarantee external validity. Secondly, clinicians often use multiple synonyms for the same disease (e.g. partial AVSD, ASD I, endocardial cushion defect, partial AV canal etc.) in medical charts. This inconsistency in nomenclature will lead to the phenomenon of “garbage in, garbage out” and it will be troublesome to achieve the reliability – compared with manual subtraction of data in a control sample - that authors were able to achieve in their study in the Royal Brompton. Considering the patient numbers needed to be able to categorize patients into large enough groups to perform the statistical analyses, it is clear that a

collaboration between multiple institutions across Europe is necessary to reach the required sample size. Importantly, different languages demand the translation of the applied algorithms, not only when it comes to coded diagnoses, but also for the performed interventions and other determinants. Herein, data standardization is essential. International efforts are ongoing to provide common data models to improve interoperability such as OMOP by OHDSI [<https://www.ohdsi.org/>].

Another challenge prior to implementing the developed algorithms across different institutions is their external validation. DL-algorithms perform extraordinary well in a single hospital setting using retrospective data. Nevertheless, in another hospital, retrospective data may differ, with patients that have another ethnical background, other co-morbidities and standard of care. Even within the same institute, algorithms may not perform as well over time due to changing standards and treatment. Furthermore, proper evaluation, similar to testing drug efficacy using randomized clinical trials, is needed to determine the added value DL-algorithms.⁶

Lastly, DL-algorithms provide useful insights, but do not directly infer causality. Therefore, caution is needed when interpreting results (e.g. frequency of hemoglobin testing does not cause bleeding). For this reason, strict regulatory approval is needed. In Europe, CE marking is necessary for AI-algorithms or clinical decision support tools according to Medical Device Directives 93/42/EEC.⁷ Nonetheless, digital errors will occur and the fundamental question is who ultimately is to be held responsible. As an explanation of results from a “black box” AI-algorithm is more difficult than traditional regression analyses, physicians and patients have to collaborate with data scientist to interpret the decisions made by the algorithm. In the end, according to GDPR, “the data subject should have the right not to be subject to a decision, which may include a measure, evaluating personal aspects relating to him or her which is based solely on automated processing”.⁸

Despite these challenges, massive opportunity lies in the near future by using the full potential of data, and thereby improve the daily care of our patients. Addressing the role of AI in our guidelines is needed in close collaboration with data scientist, regulators, ethical experts, privacy and legal experts and of course patients.

Acknowledgements:

Folkert Asselbergs is supported by UCL Hospitals NIHR Biomedical Research Centre.

References:

1. Diller et al. Machine Learning Algorithms Estimating Prognosis and Guiding Therapy in Adult Congenital Heart Disease: Data from a Single Tertiary Centre including 10,019 Patients. *Eur Heart J*. This issue.
2. Baumgartner H, Bonhoeffer P, De Groot NM, de Haan F, Deanfield JE, Galie N, Gatzoulis MA, Gohlke-Baerwolf C, Kaemmerer H, Kilner P, Meijboom F, Mulder BJ, Oechslin E, Oliver JM, Serraf A, Szatmari A, Thaulow E, Vouhe PR, Walma E; Task Force on the Management of Grown-up Congenital Heart Disease of the European Society of Cardiology (ESC); Association for European Paediatric Cardiology (AEPC); ESC Committee for Practice Guidelines (CPG). ESC Guidelines for the management of grown-up congenital heart disease (new version 2010). *Eur Heart J*. 2010;31(23):2915-57.
3. Moons P, Meijboom FJ, Baumgartner H, Trindade PT, Huyghe E, Kaemmerer H. ESC Working Group on Grown-up Congenital Heart Disease. Structure and activities of adult congenital heart disease programmes in Europe. *Eur Heart J*. 2010;31(11):1305-10.
4. Warnes CA, Liberthson R, Danielson GK, Dore A, Harris L, Hoffman JI, Somerville J, Williams RG, Webb GD. Task Force 1: the changing profile of congenital heart disease in the adult life. *J Am Coll Cardiol*. 2001;37(5):1170-
5. Franklin RCG, Béland MJ, Colan SD, Walters HL, Aiello VD, Anderson RH, Bailliard F, Boris JR, Cohen MS, Gaynor JW, Guleserian KJ, Houyel L, Jacobs ML, Juraszek AL, Krogmann ON, Kurosawa H, Lopez L, Maruszewski BJ, St Louis JD, Seslar SP, Srivastava S, Stellin G, Tchervenkov CI, Weinberg PM, Jacobs JP. Nomenclature for congenital and paediatric cardiac disease: the International Paediatric and Congenital Cardiac Code (IPCCC) and the Eleventh Iteration of the International Classification of Diseases (ICD-11). *Cardiol Young*. 2017;27(10):1872-1938.
6. Gottesman O, Johansson F, Komorowski M, Faisal A, Sontag D, Doshi-Velez F, Celi LA. Guidelines for reinforcement learning in healthcare. *Nat Med*. 2019 Jan;25(1):16-18.

7. European Commission. Guidance document – Classification of Medical Devices MEDDEV 2.4/1rev.9
<http://ec.europa.eu/DocsRoom/documents/10337/attachments/1/translations>
(18 January 2019)

8. 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)
<https://eur-lex.europa.eu/eli/reg/2016/679/oj> (18 January 2019)