

Extracting More From Less: A New Frontier for High-Throughput Clinical Phenotyping

David Ouyang MD^{1,2,*}, Susan Cheng MD MMSc MPH¹

1. Department of Cardiology, Smidt Heart Institute, Cedars-Sinai Medical Center

2. Division of Artificial Intelligence in Medicine, Cedars-Sinai Medical Center

Correspondence. David Ouyang, MD, Email david.ouyang@cshs.org.

Word count. 991

Recent innovations in computer science and information processing offer tremendous potential to improve medicine through technology – particularly when applied to electronic health records. Given that artificial intelligence and data science can uncover hidden relationships among diverse data types, the hope is that their application can uncover meaningful patterns from disparate medical information so that clinicians can more accurately detect and predict disease.^{1,2} Notwithstanding the promise and potential, there are persistent challenges posed by factors that are intrinsic to administrative databases including disparities in health access, variation in data quality, inconsistent data availability, and ambiguity in how important clinical outcomes are defined. While all of these issues serve to temper optimism in the field, the current article by Truslow et al highlights avenues for progress that can yet be made when applying sophisticated algorithms to large healthcare datasets.

Using clinical data stored in the medical records of over fifty thousand patients, Truslow et al showed that routinely acquired hematological parameters – including standard measurements from a complete blood count (CBC) assay – have strong performance for predicting major cardiovascular outcomes and mortality. As inputs into Cox proportional-hazards models, hematologic parameters demonstrated a C-index of 0.59 to 0.80 for various cardiovascular outcomes in the internal test cohort of Massachusetts General Hospital patients and a C-index of 0.60 to 0.80 in the external test cohort of Brigham and Women’s Hospital patients. Extending from the authors’ prior work,³ cardiovascular outcomes were adjudicated from discharge summary notes using a deep learning language model. The addition of age to the hematological parameters further improved model performance, consistent with the known association of aging and frailty with cardiovascular outcomes.⁴ The models predicted heart failure and all-cause mortality especially well, with concordance indices of 0.80 and 0.78, consistent with these two outcomes being among the most prevalent and most frequently seen during model training.

The importance of this study extends beyond the results alone. Perhaps foremost is the conceptual innovation of leveraging readily available yet often under-utilized diagnostic data that are routinely collected from a majority of patients. Conventional approaches to extracting new information from healthcare systems databases often require major data wrangling efforts and complex processes for bringing together disparate datasets – many of which can reside in siloed or legacy data warehouses.

Even in the absence of serving any advanced analytical purpose, substantial engineering is often needed to extract and organize medical data from different sources solely to support the clinical care operations of a hospital. Therefore, in many cases, approaches to tackling ‘big questions’ in medicine using ‘big data’ will require enormous resources and sophisticated workflows to combine variables from imaging, laboratory testing, and clinical notes. Needless to say, these are processes that extend well beyond the traditional strengths of a healthcare system. Here, Truslow et al demonstrate that it is quite possible to bypass such herculean efforts and still generate clinically relevant predictions – by gleaning information from a single frequently obtained blood test and applying natural language processing of clinical notes to scale outcomes adjudication⁵.

In addition to the conceptual innovation is a new extension of knowledge. Prior studies have shown that neutrophil-to-lymphocyte ratio values gleaned from the standard CBC are predictive of outcomes. By contrast, the focus here is on red blood cell and platelet traits. Notably, previous machine learning analyses have found that certain cardiovascular electrical⁶ and imaging⁷ features, including those obtained from routine electrocardiograms and echocardiograms, can be significantly predictive of anemia. In turn, anemia from either chronic disease or iron deficiency is known to portend poor outcomes in heart failure⁸ and been associated with mortality.⁹ Importantly, the hematologic predictors in this study represented clinical profiles that likely expand beyond anemia. While apparently specific to red cell integrity and hemostasis at the outset, the indices in this study actually also reflect the chronic functional status of multiple organs – particularly of the bone marrow, kidney, and liver – each of which can intersect with cardiovascular health and outcomes. Thus, while underscoring crosstalk between the hematologic and cardiovascular systems, Truslow et al. illustrate the potential for elucidating multiorgan system risk through analyses of highly prevalent, low-cost input data from routine CBC blood draws.

There is still more work to be done. The extent to which findings from the current study will be generalizable to other healthcare systems is not yet known. Future work is also needed to understand how the same hematologic predictors may perform similarly across different patient groups including those with or without comorbidities and prevalent disease. As with all biological markers, hematologic indices also change over time and more rapidly in some situations than in others, and so future analyses of temporal trends could offer more insights regarding the potential clinical utility of these readily available clinical measures. Moreover, the degree to which hematologic predictors of clinical risk may be responsive to therapies will need careful evaluation, particularly given that prior attempts to directly treat the anemia alone (e.g. with transfusion, erythropoietin-stimulating agents, or alternate interventions) have not produced uniformly favorable outcomes.^{9,10}

Notwithstanding the need for more work, the current study represents the next frontier of high-throughput clinical phenotyping. The greatest potential for deep learning and intelligent decision support systems to add value – earlier rather than later – is by teaching us how we can make better

use of already existing yet under-utilized medical information. Even before starting the clinical trials that will be needed to test the efficacy of intelligent decision support systems, we can still gain benefit sooner rather than later from tools that can facilitate more precise and earlier recognition of disease risk based on readily available data. Such tools can allow clinicians facing difficult decisions in clinical practice to more efficiently triage, more accurately diagnose, or more confidently manage complex clinical scenarios. Importantly, we should not shy away from the pursuit of positive change in increments. Even slight improvements in the diagnostic yield of relatively inexpensive tests – especially those that produce readily accessible data – can substantially impact outcomes at the population level.

Disclosures

None

References

1. Chen JH, Asch SM. Machine Learning and Prediction in Medicine - Beyond the Peak of Inflated Expectations. *N Engl J Med*. 2017;376:2507–2509.
2. Ouyang D, Zou J. Deep learning models to detect hidden clinical correlates. *Lancet Digit Health*. 2020;2:e334–e335.
3. Goto S, Homilius M, John JE, Truslow JG, Werdich AA, Blood AJ, Park BH, MacRae CA, Deo RC. Artificial intelligence-enabled event adjudication: estimating delayed cardiovascular effects of respiratory viruses [Internet]. *bioRxiv*. 2020; Available from: <http://medrxiv.org/lookup/doi/10.1101/2020.11.12.20230706>
4. Cheng S, Fernandes VRS, Bluemke DA, McClelland RL, Kronmal RA, Lima JAC. Age-related left ventricular remodeling and associated risk for cardiovascular outcomes: the Multi-Ethnic Study of Atherosclerosis. *Circ Cardiovasc Imaging*. 2009;2:191–198.
5. Truslow J G, Goto S, Homilius M, Mow C, Higgins J M, MacRae C A, Deo R C. Cardiovascular risk assessment using artificial intelligence-enabled event adjudication and hematologic predictors. *Circ Cardiovasc Qual Outcomes*. 2022. In Press. DOI: 10.1161/CIRCOUTCOMES.121.008007.
6. Kwon J-M, Cho Y, Jeon K-H, Cho S, Kim K-H, Baek SD, Jeung S, Park J, Oh B-H. A deep learning algorithm to detect anaemia with ECGs: a retrospective, multicentre study. *Lancet Digit Health*. 2020;2:e358–e367.
7. Hughes JW, Yuan N, He B, Ouyang J, Ebinger J, Botting P, Lee J, Theurer J, Tooley JE, Nieman K, Lungren MP, Liang DH, Schnittger I, Chen JH, Ashley EA, Cheng S, Ouyang D, Zou JY. Deep learning evaluation of biomarkers from echocardiogram videos. *EBioMedicine*. 2021;73:103613.

8. Ponikowski P, Kirwan B-A, Anker SD, McDonagh T, Dorobantu M, Drozdz J, Fabien V, Filippatos G, Göhring UM, Keren A, Khintibidze I, Kragten H, Martinez FA, Metra M, Milicic D, Nicolau JC, Ohlsson M, Parkhomenko A, Pascual-Figal DA, Ruschitzka F, Sim D, Skouri H, van der Meer P, Lewis BS, Comin-Colet J, von Haehling S, Cohen-Solal A, Danchin N, Doehner W, Dargie HJ, Motro M, Butler J, Friede T, Jensen KH, Pocock S, Jankowska EA, AFFIRM-AHF investigators. Ferric carboxymaltose for iron deficiency at discharge after acute heart failure: a multicentre, double-blind, randomised, controlled trial. *Lancet*. 2020;396:1895–1904.
9. Mazer CD, Whitlock RP, Fergusson DA, Hall J, Belley-Cote E, Connolly K, Khanykin B, Gregory AJ, de Médicis É, McGuinness S, Royse A, Carrier FM, Young PJ, Villar JC, Grocott HP, Seeberger MD, Fremes S, Lellouche F, Syed S, Byrne K, Bagshaw SM, Hwang NC, Mehta C, Painter TW, Royse C, Verma S, Hare GMT, Cohen A, Thorpe KE, Jüni P, Shehata N, TRICS Investigators and Perioperative Anesthesia Clinical Trials Group. Restrictive or Liberal Red-Cell Transfusion for Cardiac Surgery. *N Engl J Med*. 2017;377:2133–2144.
10. Holst LB, Haase N, Wetterslev J, Wernerman J, Guttormsen AB, Karlsson S, Johansson PI, Aneman A, Vang ML, Winding R, Nebrich L, Nibro HL, Rasmussen BS, Lauridsen JRM, Nielsen JS, Oldner A, Pettilä V, Cronhjort MB, Andersen LH, Pedersen UG, Reiter N, Wiis J, White JO, Russell L, Thornberg KJ, Hjortrup PB, Müller RG, Møller MH, Steensen M, Tjäder I, Kilsand K, Odeberg-Wernerman S, Sjøbø B, Bundgaard H, Thyø MA, Lodahl D, Mærkedahl R, Albeck C, Illum D, Kruse M, Winkel P, Perner A, TRISS Trial Group, Scandinavian Critical Care Trials Group. Lower versus higher hemoglobin threshold for transfusion in septic shock. *N Engl J Med*. 2014;371:1381–1391.