Automating the Study of Population Variation of Electrocardiographic Features

Isaac S. Kohane, MD, PhD

The study by Ritchie et al in this issue of Circulation uses electronic health record data and DNA biobanks to identify several genomic variants previously implicated in the variation of ECG parameters of cardiac conduction and diseases of cardiac conduction. So why is this study worthy of note?

Ever since Enthoven first named the QRS complex, investigators have sought to define what constitutes a normal complex and the diagnostic and prognostic significance of deviations from the norm. The growing understanding that there is no categorical set of normal values prompted population studies of (typically white and male) subjects numbering in the hundreds and eventually the tens of thousands. These studies did generate a more robust set of reference values and did emphasize that the notion of normal versus abnormal QRS was not appropriate and argued for “an index of the possibility of normals or abnormalities occurring at various levels” and “variations in electrocardiograms…considerably greater than the present standards would lead one to expect….” Subsequent, larger population studies including clinical trial populations with broader age and sex distributions revealed that variation in QRS characteristics in healthy individuals was larger than suspected. In parallel, several studies analyzed the clinical correlates of ECG features. For example, in 1967, Pipberger et al conducted what might today be called a “phenome scan.”

For each of the identified ECG measures, they scanned multiple constitutional features (eg, obesity) and ethnicity to assess bias and correlation. Among their findings were the significant differences in QRS measures in blacks, even when correcting for differences in the other constitutional features. Fifty years later, in the era of commodity-priced genotyping, cohort studies with tens of thousands of subjects have identified dozens of single-nucleotide polymorphisms that appear to be associated with reproducible and highly significant variations in QRS duration and several disorders of cardiac conduction, as well (eg, atrioventricular block). Several of these single-nucleotide polymorphisms implicated the SCN10A gene, a subunit of one of the voltage-gated sodium channels, also found in the study of Ritchie et al.

Also, over the past 3 decades, with the deployment of electronic health records (EHRs), informaticians at leading healthcare systems demonstrated how ECG data could be integrated with other clinical data obtained in the course of healthcare delivery and used to predict outcomes, such as mortality, in the very same populations being cared for. These early efforts laid the foundations for exploiting the low incremental costs of using EHR data to rapidly characterize and select study populations. As phenotyping and sample acquisition became the major costs in disease genomics studies, the use of EHRs to create an instrumented health enterprise for genetic discovery research by the use of the informational byproducts of healthcare delivery (ie, clinical documentation) and landed or discarded clinical blood samples has become increasingly attractive. Multiple studies have shown that this EHR-driven approach is feasible, accurate, and cost-effective, and several national funding agencies now support these studies internationally.

In this context, the contribution of the study by Ritchie et al is 2-fold. First is the demonstration that EHR-driven phenotyping can be used to accurately select patients and reproduce genomic associations (principally pointing to the genes SCN5A and SCN10A for conduction disorders in a manner that scales cost-effectively to much larger population studies. Specifically, the EHRs were used to identify healthy individuals, to quantify their ECG measures, and to select the corresponding genotypes for the same individuals. Second, the insights provided by the hypothesis-free inversion of conventional genome-wide studies. That is, the investigators selected the most significant single-nucleotide polymorphisms (with respect to QRS variation) and scanned the entirety of the diagnoses of all patients in the EHR to determine which diagnoses were significantly correlated with those common genomic variants. These included a variety of cardiac arrhythmias. Moreover, they used the EHR to longitudinally track the patients originally identified as healthy in their QRS study, and they found that 3% of patients developed atrial fibrillation or atrial flutter at some point at least 1 month following the normal ECG, and 11% were coded as having a variety of subsequent arrhythmias. This in silico cohort selection and longitudinal study is in many ways a model case of precision medicine as defined in a recent Institute of Medicine report. That report argued for the creation of an “information commons” with multiple layers of measurement all linked to individual patients to accelerate the acquisition of biomedical knowledge. The report also emphasized the importance of population studies that include the full complexity of our patient populations, including their ethnic heterogeneity, polypharmacy, and comorbidities. The report also anticipated redrawing the current categorical diagnostic
or disease boundaries as multidimensional and probabilistic measures that draw directly from the quantitative measures available in the information commons. The work of Ritchie et al provides additional evidence of the feasibility and efficacy of the precision medicine model.

Several important loose ends remain in this study. For example, members of underrepresented minorities were specifically excluded, even though there is at least a 50-year history of ethnicity-specific variation in ECG characteristics. Because these same underrepresented minorities are often overrepresented in academic health centers, the same EHR-driven approach could be readily and rapidly used to study the genomic basis of those minorities, even though there is at least a 50-year history of ethnicity-specific variation in ECG characteristics. Because these underrepresented minorities are often overrepresented in academic health centers, the same EHR-driven approach could be readily and rapidly used to study the genomic basis of those differences. Also, this study relied heavily on billing codes rather than the fine-grained diagnostic assessment of clinicians. The systematic application of natural language-processing techniques to codify the content of clinical notes in EHRs will minimize the biases and lack of granularity that come from the use of billing data. Most ambitiously, restructuring the phenome scan to include broader processes such as inflammation or thrombosis may help speed the genomic characterization of the endophatypes that underlie multiple diseases.

**Sources of funding**

Dr Kohane is partially supported by National Institutes of Health National Center for Biomedical Computing grant SU54LM008748

**Disclosures**

Dr Kohane is funded by the National Institutes of Health eMERGE network that also funded the work of Ritchie et al (under a separate grant).

**References**


**Keywords**: Editorsials — arrhythmia (heart rhythm disorders) — biobanks — bioinformatics — electronic health records — medical informatics
Automating the Study of Population Variation of Electrocardiographic Features
Isaac S. Kohane

Circulation. 2013;127:1357-1358; originally published online March 5, 2013; 
doi: 10.1161/CIRCULATIONAHA.113.001852
Circulation is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2013 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7322. Online ISSN: 1524-4539

The online version of this article, along with updated information and services, is located on the
World Wide Web at:
http://circ.ahajournals.org/content/127/13/1357

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published
in Circulation can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial
Office. Once the online version of the published article for which permission is being requested is located,
click Request Permissions in the middle column of the Web page under Services. Further information about
this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Circulation is online at:
http://circ.ahajournals.org//subscriptions/