“Retrofitting” established genetic disorders and diseases through big data and phenomics

William D. Graf, MD, and Robert J. Shprintzen, PhD

Neurology: Clinical Practice October 2020 vol. 10 no. 5 375-376 doi:10.1212/CPJ.0000000000000784

Correspondence

Dr. Graf
wgraf@connecticutchildrens.org

“The value of experience is not in seeing much, but in seeing wisely”—Sir William Osler.

Throughout the history and evolution of medical diagnostics, astute clinicians used “gestalt” (or pattern recognition) methods to identify and describe specific neurologic disorders, neurodegenerative diseases, and genetic syndromes. The application of innovative technologies allowed valid clinical observations to be corroborated by various biomarkers, such as laboratory measurements and histopathology, and ultimately by DNA sequencing. In the current era of “omics” sciences (e.g., genomics, proteomics, or metabolomics), advances in bioinformatics and large health care databases demonstrate the potential to further expand diagnostic and therapeutic knowledge. At the organism level, phenomics is the study of phenotypic changes in response to genetic variants and environmental influences.¹

In this issue of *Neurology: Clinical Practice*, Bonkowsky et al.² retrospectively interrogated a large clinical database (Pediatric Health Information System [PHIS]) using International Classification of Diseases, Clinical Modification (ICD-9-CM or ICD-10-CM) codes for 4 genetically defined childhood onset leukodystrophies (metachromatic leukodystrophy [E75.25], X-linked adrenoleukodystrophy [E71.52x], Krabbe disease [E75.23], and Hurler disease [E76.01]). This national electronic health records (EHRs)-based phenome-wide association study (PheWAS) analyzed the association of comorbidities in patients with leukodystrophy compared with age-matched controls. The aim of this leukodystrophy PheWAS was to characterize disease burden and improve clinical care. The study identified various “phecode” associations that occurred more frequently in the leukodystrophy diagnosis groups. Some of the unexpected study findings may be useful, but other findings require methodological interpretation.

Clinical medicine and health care are increasingly becoming “big data”-intensive enterprises. Clinical application of PheWAS, genome-wide association studies (GWASs), and artificial intelligence (AI) technologies are used in analyzing vast quantities of clinical data from various databases, including the EHR.³ Whereas GWAS begins with a certain phenotype and systematically analyzes variants across the entire genome in association with the phenotype, a PheWAS begins with a certain genetic variant and systematically analyzes any associated phenotypes.⁴ GWAS can identify several genetic associations to a phenotype with polygenic traits. The PheWAS can identify pleiotropy (i.e., the expression of multiple traits by a single gene variant). AI in health care employs pathfinding algorithms and software to emulate human cognition in the analysis of complex medical data, including patient data from multi-omic approaches and clinical, behavioral, environmental, and pharmacologic data stored in large databases or published in the medical literature.⁵

PheWAS and GWAS face certain challenges in their use of EHR-based data. First, linking most “omics” data with phenotypes is difficult because omics volume is enormous compared with the relatively small number of clinical phenotypes—at least when broad descriptive categorical diagnoses (e.g., “constipation,” “fatigue,” and “epilepsy”) and grouped ICD-CM codes (or phecodes) are used. Second, whereas genomic data are static data that do not change over time,

RELATED ARTICLE

Research

Novel and known morbidities of leukodystrophies identified using phenome-wide association study

[Page 406](#)

Connecticut Children's (WDG), Farmington; and The Virtual Center for Velo-Cardio-Facial Syndrome (RJS), Manlius, NY.

Funding information and disclosures are provided at the end of the article. Full disclosure form information provided by the authors is available with the full text of this article at Neurology.org/cp.

phenomic data are dynamic data that will change after they are recorded. For PheWAS to achieve clinical utility, clinicians and data entry billing administrators must continuously update EHR data to guarantee their accuracy. Unless temporary, “noisy,” and extraneous ICD-CM codes are “resolved” (or purged) from EHR databases, the PheWAS will be continuously plagued by “garbage in, garbage out” and uncertainty.⁶ Most importantly, because disease phenotypes evolve during their clinical course, the disease stage at the time of the PheWAS must be correlated with real-time or “temporal” data.⁷

For example, Bonkowsky et al.² found a high association of the categorical diagnosis “developmental delay” in patients with specific leukodystrophies. In clinical practice, most child neurologists accept the term “developmental delay” (F80.9) only as a referral diagnosis or a “chief concern.” Child neurologists and other perceptive clinicians can distinguish milder developmental concerns (e.g., R62.50 or F82) from a true “neurologic disorder” (G96.9) or a “neurodegenerative disease” (G31.9) such as leukodystrophies, which are progressive, metabolic, genetic diseases that variably affect the brain, spinal cord, and peripheral nerves. Moreover, in this study, if only 54% of the patients in the leukodystrophy PheWAS had phecodes for “developmental delay,” then conversely 46% of these patients did not have “delay,” which further obfuscates the meaning of this EHR ICD-10-CM finding. Clearly, this PHIS diagnostic association is present but unhelpful.

Likewise, clinical experience will reason that biology alone is insufficient to understand disease.⁸ Disorders and diseases are expressed in patients and diagnosed by doctors through a process that is influenced by training, experience, and environmental circumstances.⁹ The leukodystrophy PheWAS considered “constipation” (K59*) to be an unexpected association in the nontransplant leukodystrophy cohort and proposed “proactive treatment of constipation” as a potentially useful outcome of the PheWAS. However, “constipation” may represent a billable ICD-10-CM code for the ongoing treatment of a self-evident, secondary complication of severe disability (i.e., not necessarily an essential biological element of a genetic disease).

Physicians from different eras looking at the same patient may interpret signs and symptoms differently.⁹ The clinical skills acquired by older physicians may be less tangible for younger

physicians who are trained during the EHR era of mandatory time-consuming data entry. Clinical guidelines and treatment algorithms will evolve—certainly through the contributions of PheWAS data—but the care of individual patients will remain the basis of medical practice.⁹ Thus, the integration of big data into future health care may be an attempt to “retrofit” current medical practice. It is still uncertain whether the transformative rewards of AI technologies will enhance medical practice or merely be used as a justification to hasten clinical encounters and increase physician “productivity.” Whether AI will allow physicians more time to devote to their patients and to practice the art of (personalized) medicine remains the biggest question.¹⁰

Author contributions

W.D. Graf and R.J. Shprintzen: Drafting/revising the manuscript.

Study funding

No targeted funding reported.

Disclosure

The authors report no disclosures relevant to the manuscript. Full disclosure form information provided by the authors is available with the full text of this article at Neurology.org/cp.

References

1. Bilder RM, Sabb FW, Cannon TD, et al. Phenomics: the systematic study of phenotypes on a genome-wide scale. *Neuroscience* 2009;164:30–42.
2. Bonkowsky JL, Wilkes J, Ying J, Wei WQ. Novel and known morbidities of leukodystrophies identified using phenome-wide association study. *Neurol Clin Pract* 2019;10:406–414.
3. Verma A, Ritchie MD. Current scope and challenges in phenome-wide association studies. *Curr Epidemiol Rep* 2017;4:321–329.
4. Robinson JR, Denny JC, Roden DM, Van Driest SL. Genome-wide and phenome-wide approaches to understand variable drug actions in electronic health records. *Clin Transl Sci* 2018;11:112–122.
5. Wang F, Preiner A. AI in health: state of the art, challenges, and future directions. *Yearb Med Inform* 2019;28:16–26.
6. Sturmberg JP. Evidence-based medicine-not a panacea for the problems of a complex adaptive world. *J Eval Clin Pract* 2019;25:706–716.
7. Wu PY, Cheng CW, Kaddi CD, Venugopalan J, Hoffman R, Wang MD. -Omic and electronic health record big data analytics for precision medicine. *IEEE Trans Biomed Eng* 2017;64:263–273.
8. Visscher PM, Wray NR, Zhang Q, et al. 10 years of GWAS discovery: biology, function, and translation. *Am J Hum Genet* 2017;101:5–22.
9. Holzinger A, Dehmer M, Jurisica I. Knowledge discovery and interactive data mining in bioinformatics—State-of-the-Art, future challenges and research directions. *BMC Bioinformatics* 2014;15(suppl 6):i11.
10. Truog RD. Of slide rules and stethoscopes: AI and the future of doctoring. *Hastings Cent Rep* 2019;49:3.