

## CARE INNOVATIONS

# Implementation of a Machine-Learning Algorithm in the Electronic Health Record for Targeted Screening for Familial Hypercholesterolemia

## A Quality Improvement Study

### GOALS AND VISION OF THE PROGRAM

Familial hypercholesterolemia (FH) is a lipid disorder that results in elevated serum LDL (low-density lipoprotein) cholesterol and markedly increased cardiovascular risk.<sup>1,2</sup> Classical observational data suggest that prevalence of heterozygous FH is  $\approx$ 1:250, and it is estimated that only 10% of patients with FH in the United States have been diagnosed.<sup>1,2</sup> Early and timely diagnosis of FH reduces cardiovascular risk, which heightens the need for targeted screening.<sup>2,3</sup> To increase the detection rates of FH, several population and targeted screening strategies have been recommended and implemented. For example, mass genetic testing in the workplace and cascade genetic screening have been used in a few settings in the United States. Machine learning models trained on electronic medical record (EMR) data represent another promising approach to identify high-risk populations enriched with FH patients, but deployment of machine learning algorithms in cardiovascular medicine has been a historically challenging process.<sup>4</sup>

Recently, The Familial Hypercholesterolemia Foundation developed the flag, identify, network, deliver FH (FIND FH) machine learning algorithm to identify yet-to-be diagnosed FH within millions of individual EMRs.<sup>5</sup> FIND FH is a random forest-based algorithm trained on deidentified, structured EMR data from 939 individuals who were diagnosed with FH at specialty lipid clinics. The model selects 75 features ranging from patient demographics to prescriptions and laboratory data to predict the probability of a patient having FH. In the original study, the model demonstrated robust performance in predicting patients with higher risk of FH at both national (170 416 201 patients) and single health care system (173 733 patients from Oregon Health & Science University) levels, identifying 87% and 77% of patients in 2 independent cohorts as having a high enough suspicion of FH to warrant further evaluation and treatment (likely FH).<sup>5</sup> At the University of Pennsylvania Healthcare System (UPHS), an internal validation of 414 patients flagged by FIND FH revealed that 29% of patients with FIND FH score  $>0.2$  had probable or definite FH (unpublished data). However, no prior studies involving FIND FH had developed an implementation framework, complete with an outreach process, to integrate the algorithm into clinical care.

The purpose of this study was to implement an observational trial of a HIPAA-compliant, IRB-approved screening and outreach program based on FIND FH as a case study of how machine learning algorithms could be deployed and utilized in a large health care system. Through this initiative, we assessed (1) the diagnostic rate of FH among clinical results from patients flagged by the algorithm, (2) the treatment of flagged patients in a preventive cardiology setting, and (3) barriers in implementation of the algorithm to inform future quality improvement initiatives.

Samip Sheth, MB  
Paul Lee, MS  
Archna Bajaj, MD  
Marina Cuchel<sup>ORCID</sup>, MD PhD  
Jihane Hajj, DrNP  
Daniel E. Soffer, MD  
Gayley Webb, CrNP  
Erik Hossain, BS  
Yulia Borovskiy, MS  
Marjorie Risman, MA  
Kelly D. Myers<sup>ORCID</sup>, BS  
Katherine A. Wilemon<sup>ORCID</sup>, BS  
Daniel J. Rader<sup>ORCID</sup>, MD  
Douglas Jacoby<sup>ORCID</sup>, MD

**Key Words:** genetic testing  
■ hyperlipoproteinemia type II ■ mass screening ■ quality improvement ■ risk factors

© 2021 American Heart Association, Inc.

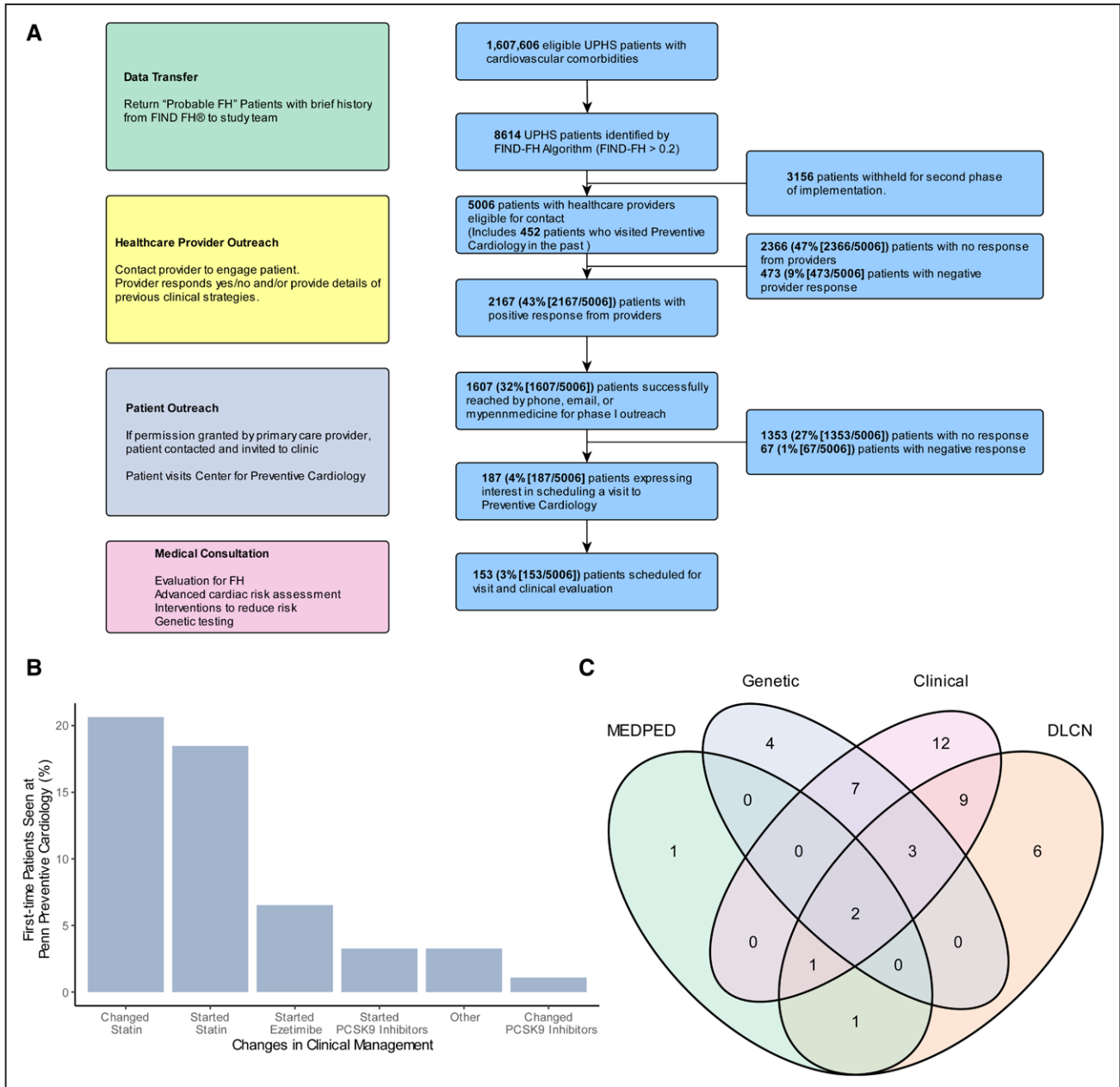
<https://www.ahajournals.org/journal/circoutcomes>

## DESIGN AND IMPLEMENTATION OF THE INITIATIVE

The Center for Preventive Cardiology and Lipid Management at UPHS coordinated the design, implementation, and assessment of the initiative. The staff at the center include cardiologists, lipidologists, geneticists, nurse practitioners, a dietician, and nursing and administrative staff. The target population within the health care system were those patients who were identified by the algorithm to be at elevated risk for FH (FIND FH

score, >0.2). To mimic a routine referral process as a part of standard clinical care, the center implemented an outreach model that involved both the prospective patients and their primary care providers (PCPs).

From August 2018 to August 2019, a trained clinical research coordinator contacted the PCP for each patient flagged by the algorithm through a clinical letter or staff message sent in the EPIC EMR system (Epic Systems Corporation, Verona, WI), by phone call or by email asking the patient to be referred to the Center for Preventive Cardiology (Figure [A]). Patients >75



**Figure.** Implementation of flag, identify, network, deliver familial hypercholesterolemia (FIND FH) in the University of Pennsylvania Healthcare System (UPHS).

**A**, HIPAA-compliant recruitment model for implementation of FIND FH. **B**, Changes in clinical management among 92 patients who visited Penn Preventive Cardiology for the first time as a part of the initiative. **C**, Distribution of diagnostic method among 46 patients diagnosed with familial hypercholesterolemia at Penn Preventive Cardiology. DLCN indicates Dutch Lipid Clinic Network; DLCNS, Dutch Lipid Clinic Network Score; HIPAA, the Health Insurance Portability and Accountability Act of 1996; and MEDPED, Make Early Diagnosis to Prevent Early Death; and PCSK9 inhibitor, proprotein convertase subtilisin/kexin type 9 inhibitor.

years of age and those without contact information were not contacted. If a provider declined permission or did not respond after one follow-up request, their patients were not further contacted. If the patient had visited a provider of the Center for Preventive Cardiology and Lipid Management in the past, they were automatically deemed eligible for contact. The initiative was approved by the Institutional Review Board of the University of Pennsylvania.

Following the provider's approval, eligible patients were contacted directly by a message sent to a patient portal (MyPennMedicine), by phone call or by email. The message included information about FH, the purpose of the quality improvement initiative, and an offer to schedule a visit with a provider in the preventive cardiology clinic. If the patient did not respond back, another follow-up message was sent 1 week after the initial message. Before each appointment, the provider was notified that a patient had been scheduled through the initiative; however, the provider was blinded to the FIND FH score as to not bias the diagnosis. The clinical evaluation by the provider followed the standard of care in our preventive cardiology program. As all the participants were deemed to be at elevated risk for FH, all participating providers agreed to offer genetic testing of the *LDLR*, *APOB*, and *PCSK9* genes to all participants to reach a definitive diagnosis. While the consultation visit was billed to the patient, the genetic testing was offered free of charge and completed via next-generation sequencing with microarray confirmation at Quest Diagnostics. The Center for Preventive Cardiology has historically offered genetic testing as a part of standard clinical care, but patients typically have a copayment that depends on their insurance.

After the initial visit, all patient data were recorded into the EMR, and all patients were notified of their genetic testing result by a phone call or an email from a trained clinical research coordinator. Starting September 2019, the clinical data from the initial visits were abstracted retrospectively by a clinical research coordinator. All patients were rescored following the Dutch Lipid Clinic Network criteria and the Make Early Diagnosis to Prevent Early Death criteria. Scoring was based on available clinical data and genetic test results from the initial visit.<sup>3</sup> Any patient who was not diagnosed with FH during the clinical visit but (1) possessed a pathogenic, likely pathogenic, or variant of unknown significance in the three genes tested or (2) had probable FH based on the diagnostic criteria was asked to be reevaluated by the physician.

## RESULTS OF THE INITIATIVE

Among 1607606 eligible patients with cardiovascular comorbidities in the UPHS EMR, 8614 were flagged as having a FIND FH score  $>0.2$ , indicating likely FH. We attempted to contact the health care providers for 5006 of these patients (442 health care providers; Figure [A])

whose individual identities were provided by LabCorp (a HIPAA covered entity) for the first phase of the implementation study defined by the scheduling capacities of the administrative staff. Identities and FIND FH scores of a randomly selected subset of 3614 patients were withheld to be provided in the second phase of the project. Of the 442 contacted PCPs, 223 (53%) responded. These 223 providers were associated with a total of 2640 patients, of whom 2167 (43% [2167 of 5006]) remained eligible to contact and 473 (9% [473 of 5006]) were ineligible to contact after a favorable or an unfavorable response by their providers, respectively. Of these 2167 patients, 1607 (32% [1607 of 5006]) were successfully contacted through phone, email, or MyPennMedicine message. Of the patients contacted, 187 (4% [187 of 5006]) expressed interest in participation, and 153 (3% [153 of 5006]) were ultimately seen in the preventive cardiology clinic.

The final participant population demonstrated characteristics consistent with an at-risk population for FH. The median LDL cholesterol of participants at enrollment was 151 (interquartile range, 128–180) mg/dL, and the median total cholesterol was 234 (interquartile range, 207–265) mg/dL. Of the 153 patients, 68 (44.4%) had a first-degree relative with known premature atherosclerotic cardiovascular disease, 8 (5.2%) presented with tendinous xanthomas, and 14 (9.2%) presented with corneal arcus. Based on self-reports, 92 (60%) had never visited a clinical lipid specialist in the UPHS before the initiative, and 112 (73%) of 153 patients consented to and received genetic testing for FH for the first time as a part of the initiative.

Following the initial visit, 46 patients were ultimately diagnosed with FH by (1) phenotypic clinical assessment by a physician or (2) Dutch Lipid Clinic Network/Make Early Diagnosis to Prevent Early Death criteria or (3) the presence of an FH mutation (Figure [C]). Of the 153 seen in the clinic, 31 were newly diagnosed patients who visited the specialty lipid clinic for the first time, while 15 patients had visited the clinic previously and were either already diagnosed with FH or reevaluated for FH and received intensification of care. A total of 16 patients tested positive for FH by genetic testing (14% [16 of 112]), and 42 patients received a diagnosis of FH based on clinical assessment or diagnostic criteria. Of the patients who received molecular diagnoses for FH, we observed a wide range of mutations: 7 possessed functional variants in *LDLR*, 3 possessed functional variants in *PCSK9*, and 6 possessed functional variants in *APOB*. Using the Dutch Lipid Clinic Network or the Make Early Diagnosis to Prevent Early Death criteria, only 23 (50% [23 of 46]) of genetically confirmed patients would have been classified as having possible FH (Figure [C]). Of the 46 diagnosed patients, the median LDL cholesterol was 196 (interquartile range, 151.25–217.5) mg/dL and the median total cholesterol was 275.5 (interquartile range, 233.25–301.25) mg/dL.

We then determined the impact of the single consultation visit in improving the clinical management of the participants. While less than half of the patients were diagnosed with FH, most of the patients in the initiative saw changes in their clinical management regardless of their final diagnoses. Among the 92 patients who visited the preventive cardiology clinic for the first time, 9 (10%) had an LDL cholesterol level exceeding 190 mg/dL and 49 (53%) underwent changes in clinical management. The most common changes were intensification (39% [19 of 49]) or initiation (35% [17 of 49]) of statin regimen (Figure [B]). If the patient tested genetically positive, they received further consultation regarding cascade screening for family members.

## LOCAL CHALLENGES IN IMPLEMENTATION

There were several challenges in the implementation of the algorithm unique and nonspecific to this initiative. First, while we report the diagnosis rates and clinical characteristics of flagged patients who were evaluated at the clinic, based on our current cohort of patients, we cannot draw any robust conclusions about model performance in detecting FH without a control arm consisting of patients not flagged by the algorithm (FIND FH probability, <0.2). However, another prospective study at UPHS (IN TANDEM [Integrating Active Case-finding With Next-generation Sequencing for Diagnosis Through Electronic Medical Records]; <https://www.clinicaltrials.gov>; unique identifier: NCT03253432) validating the performance of FIND FH in predicting FH patients at different risk scores is currently in progress. Second, while provider participation was moderately high (53%), patient participation was low despite the offer of genetic testing free of charge. Patients who refused scheduling provided several explanations for doing so, including lack of awareness of FH and cardiovascular prevention, method of contact from the Preventive Cardiology office instead of PCP, long drive to visit the clinic from hospitals and clinics in other regions and states, inability to afford billed visit or parking, and wanting to directly followup with a PCP. Lastly, while patients with an *International Classification of Diseases* code for FH (E78.01) were excluded, the inconsistency in the usage of the code resulted in the algorithm flagging several known FH patients. Given that the *International Classification of Diseases, Tenth Revision* code for FH was approved by the Centers for Medicare and Medicaid Services in 2016, the limited usage of this *International Classification of Diseases* code was understandable.

## TRANSLATION TO OTHER SETTINGS

We need innovation in improving response rates if the FIND FH algorithm and other EMR-based tools are to

have a meaningful impact. Namely, the low yield (3% [153 of 5006]) of patients who were ultimately seen in the clinic suggests that there is high resistance in both getting the patient referred to a secondary clinic by a provider and convincing the patient to schedule a visit after the referral. In this study, the majority of the flagged patients (57%) were deemed ineligible for contact due to lack of or negative provider response. In other settings, the specialty clinic could consider prioritizing increased buy-in and coordination with super-FH providers or PCPs who see a high number of patients flagged by the FIND FH algorithm. Providing education and data about the performance of the algorithm to all PCPs in the program before outreach may also enhance awareness and subsequent care for FH in primary care settings. In addition, prioritizing providers whose flagged patients have on average higher FIND FH probabilities may increase successful outreach. Overall, we expect that the greatest opportunity for improving efficacy of outreach lies in primary care settings, and future work will focus on creating an outreach model that encourages PCPs to provide more formal referrals to the clinic instead of direct electronic messages to the patients from the clinic. One way to achieve this may be to establish staff at high-volume primary care sites to discuss the initiative with patients and providers who may be interested. An alternate EMR-based notification model where a provider receives an electronic notification during a scheduled encounter with a flagged patient could also encourage referrals. Implementations in other settings should also carefully consider how patients may perceive the value proposition of the cost and time associated with scheduling a new outpatient visit for FH workup. For example, the first phase of this study was completed before the outbreak of coronavirus disease 2019 (COVID-19) and before virtual medical visits were implemented at the Center for Preventive Cardiology; the implementation may have seen much greater uptake if virtual visitation was made available.

## SUMMARY AND FUTURE DIRECTIONS

In a large health care system, we prospectively implemented a random forest algorithm, FIND FH, through a quality improvement initiative to perform targeted screening for FH. The algorithm identified a population of patients at an elevated risk for FH, who were invited to be evaluated for FH through our outreach program. While a low percentage of eligible patients ultimately scheduled a visit to our Center for Preventive Cardiology, among those patients who were evaluated, 30% were diagnosed with FH. Most of the first-time patients who visited the clinic saw modification in lipid-lowering treatments, mostly intensification or initiation of statin therapy.

In summary, our implementation of FIND FH in UPHS demonstrates well-recognized challenges in scaling and

increasing the utility of machine learning algorithms for screening diseases with a range of phenotypes.<sup>4</sup> Future directions for the implementation study include working closely with PCPs before future implementation to discuss potential concerns about FIND FH, convening focus groups of patients and physicians to understand how they would like to be approached and what would motivate them to participate in the initiative, and having the initial invitation to the patients come from both the PCP and Preventive Cardiology Center with the option for virtual or in-person visits. Our study ultimately points out the need for better dissemination and implementation studies of highly specialized machine learning tools in a rare disease prevention setting.

## ARTICLE INFORMATION

### Correspondence

Douglas Jacoby, MD, University of Pennsylvania, Pennsylvania Hospital, Farm Journal Bldg, 3rd Floor, 230 W Washington Sq Philadelphia, PA 19106. Email [douglas.jacoby@penmedicine.upenn.edu](mailto:douglas.jacoby@penmedicine.upenn.edu)

### Affiliations

Department of Medicine (S.S., P.L., A.B., M.C., J.H., D.S., G.W., M.R., D.R., D.J.) and Data Analytics Center (E.H., Y.B.), Perelman School of Medicine at the University of Pennsylvania, Philadelphia. The Familial Hypercholesterolemia Foundation, Pasadena, CA (K.M., K.W.).

### Acknowledgments

This study was supported by The Familial Hypercholesterolemia Foundation that developed the FIND FH machine learning algorithm. We would like to thank William Howard, Mary McGowan, Amanda Sheldon, Dave Staszak, Cynthia Mays, Kylie Boehler, Jeff Radcliff, Carrie Castonguay, and David Wrenn for their

feedback on the manuscript. We would also like to thank the Penn Data Analytics Center for their assistance in assembling the information used in this study.

### Sources of Funding

This quality improvement initiative was funded by The Familial Hypercholesterolemia Foundation. Funding for genetic testing was provided by Quest Diagnostics. The FIND FH program was funded in part by Amgen.

### Disclosures

None.

## REFERENCES

1. Abul-Husn NS, Manickam K, Jones LK, Wright EA, Hartzel DN, Gonzaga-Jauregui C, O'Dushlaine C, Leader JB, Lester Kirchner H, Lindbuchler DM, et al. Genetic identification of familial hypercholesterolemia within a single U.S. health care system. *Science*. 2016;354:aaf7000. doi: 10.1126/science.aaf7000
2. Gidding SS, Champagne MA, de Ferranti SD, Defesche J, Ito MK, Knowles JW, McCrindle B, Raal F, Rader D, Santos RD, et al; American Heart Association Atherosclerosis, Hypertension, and Obesity in Young Committee of Council on Cardiovascular Disease in Young, Council on Cardiovascular and Stroke Nursing, Council on Functional Genomics and Translational Biology, and Council on Lifestyle and Cardiometabolic Health. The agenda for familial hypercholesterolemia: a scientific statement from the American Heart Association. *Circulation*. 2015;132:2167–2192. doi: 10.1161/CIR.0000000000000297
3. McGowan MP, Hosseini Dehkordi SH, Moriarty PM, Duell PB. Diagnosis and treatment of heterozygous familial hypercholesterolemia. *J Am Heart Assoc*. 2019;8:e013225. doi: 10.1161/JAHA.119.013225
4. Safarova MS, Liu H, Kullo IJ. Rapid identification of familial hypercholesterolemia from electronic health records: the SEARCH study. *J Clin Lipidol*. 2016;10:1230–1239. doi: 10.1016/j.jacl.2016.08.001
5. Myers KD, Knowles JW, Staszak D, Shapiro MD, Howard W, Yadava M, Zuzick D, Williamson L, Shah NH, Banda JM, et al. Precision screening for familial hypercholesterolemia: a machine learning study applied to electronic health encounter data. *Lancet Digital Health*. 2019;1:e393–e402. doi: 10.1016/S2589-7500(19)30150-5