AMIA
INFORMATICS PROFESSIONALS. LEADING THE WAY.

OXFORD

Research and Applications

Precision population analytics: population management at the point-of-care

Paul C. Tang,[1] Sarah Miller,[2] Harry Stavropoulos,[3] Uri Kartoun,[3] John Zambrano,[4] and Kenney Ng[3]

[1]Stanford Clinical Excellence Research Center, Stanford University, Stanford, California, USA, [2]IBM Research, Cambridge, Massachusetts, USA, [3]Center for Computational Health, IBM Research, Cambridge, Massachusetts, USA, and [4]Atrius Health Academic Institute, Atrius Health, Boston, Massachusetts, USA

Corresponding Author: Paul C. Tang, MD, Clinical Excellence Research Center, Department of Medicine, Gunn-SIEPR Building, 366 Galvez St, Stanford, CA 94305, USA (paultang@stanford.edu)

ABSTRACT

Objective: To present clinicians at the point-of-care with real-world data on the effectiveness of various treatment options in a precision cohort of patients closely matched to the index patient.

Materials and Methods: We developed disease-specific, machine-learning, patient-similarity models for hypertension (HTN), type II diabetes mellitus (T2DM), and hyperlipidemia (HL) using data on approximately 2.5 million patients in a large medical group practice. For each identified decision point, an encounter during which the patient's condition was not controlled, we compared the actual outcome of the treatment decision administered to that of the best-achieved outcome for similar patients in similar clinical situations.

Results: For the majority of decision points (66.8%, 59.0%, and 83.5% for HTN, T2DM, and HL, respectively), there were alternative treatment options administered to patients in the precision cohort that resulted in a significantly increased proportion of patients under control than the treatment option chosen for the index patient. The expected percentage of patients whose condition would have been controlled if the best-practice treatment option had been chosen would have been better than the actual percentage by: 36% (65.1% vs 48.0%, HTN), 68% (37.7% vs 22.5%, T2DM), and 138% (75.3% vs 31.7%, HL).

Conclusion: Clinical guidelines are primarily based on the results of randomized controlled trials, which apply to a homogeneous subject population. Providing the effectiveness of various treatment options used in a precision cohort of patients similar to the index patient can provide complementary information to tailor guideline recommendations for individual patients and potentially improve outcomes.

Key words: clinical decision support, machine learning, population health management, electronic health records, treatment outcome

INTRODUCTION

Practicing evidence-based medicine is a central tenet of care delivery and population health management. Questions have been raised, however, about the applicability of generalized guidelines to an individual patient[1] and whether individualized guidelines could lead to better care and

lower costs.[2] Guidelines are often referred to as evidence-based when derived from results of randomized controlled clinical trials (RCTs) or consensus expert opinion. Unfortunately, RCTs cover only a minority of clinical conditions and typically involve a relatively small number of study subjects. More importantly, by design, RCTs attempt to remove

as many confounders as possible in order to attribute the results to the randomly assigned intervention under study. Consequently, the inclusion and exclusion criteria can be very stringent. As a result, the subjects under study represent a homogeneous population that is not representative of the community-at-large. Community patients are often older, include more women, have more comorbidities, and have behavioral-activation states that may be different from subjects who were recruited for and consented to participate in the controlled setting of a clinical trial.[3] For example, the SPRINT clinical trial[4] is a large multicenter RCT frequently cited as a basis for evidence-based hypertension guidelines. When the inclusion and exclusion criteria used to recruit patients for the SPRINT trial were applied to the community population of patients with hypertension (HTN) at the study site, only 5.6% of ~129K patients with hypertension would have qualified to participate in the trial. Furthermore, only a small self-selected subset of those eligible to participate in a clinical trial actually enroll.[5] Even within the homogeneous clinical-trial population, there are notable heterogeneous treatment effects that are masked by the fact that trial results only compare the mean values of the populations studied.[6]

Ideally, physicians making decisions at the point-of-care should have available not only guidance from RCTs, where available and applicable, but also information about the lessons learned from everyday practice that are captured as observational data in electronic health records (EHRs). The question is which subsets of patients are most relevant to a specific patient. Guidelines are commonly stratified by demographic variables such as age, gender, and race. Although it is commonly thought that EHRs contain only clinical data, some EHR data can also be used to infer behavioral activation status (eg, timely completion of health maintenance procedures, medication adherence) and other determinants of health, including social determinants.[7,8]

We used machine learning methods to develop disease-specific patient-similarity models to create *precision cohorts* for an index patient. We define a precision cohort as a set of patients closely matched both to the characteristics of the index patient and to the precise clinical situation (eg, the control status of the clinical condition, active medications, clinical and behavioral characteristics of the individual). Within each precision cohort, one can examine the clinical responses of similar patients to various treatment options administered by other physicians in the same organization, a significant influencer in behavioral change management. Consequently, it is possible to quantitatively characterize best-outcomes treatment options (treatment options that produce the highest level of disease control) in the organization for patients similar to the index patient in similar clinical situations. We refer to the empirically observed best-outcomes practices as best practices. The best practices can be visualized and presented to the physicians at the point-of-care within the EHR.

In this paper, we describe the methods used to develop the models and to present results of the precision-cohort analysis at the point-of-care in a large medical group practice.

MATERIALS AND METHODS

Study data
Deidentified EHR data from a large health care provider in eastern Massachusetts, containing over 20 years of longitudinal data for approximately 2.5 million patients were used in this study. Details of the EHR data are provided in Supplementary Section S1. The modeling and analysis focused on 3 common chronic disease conditions: hypertension (HTN), type 2 diabetes mellitus (T2DM), and hyperlipidemia (HL). The Western Institutional Review Board reviewed the study and granted a waiver of HIPAA authorization.
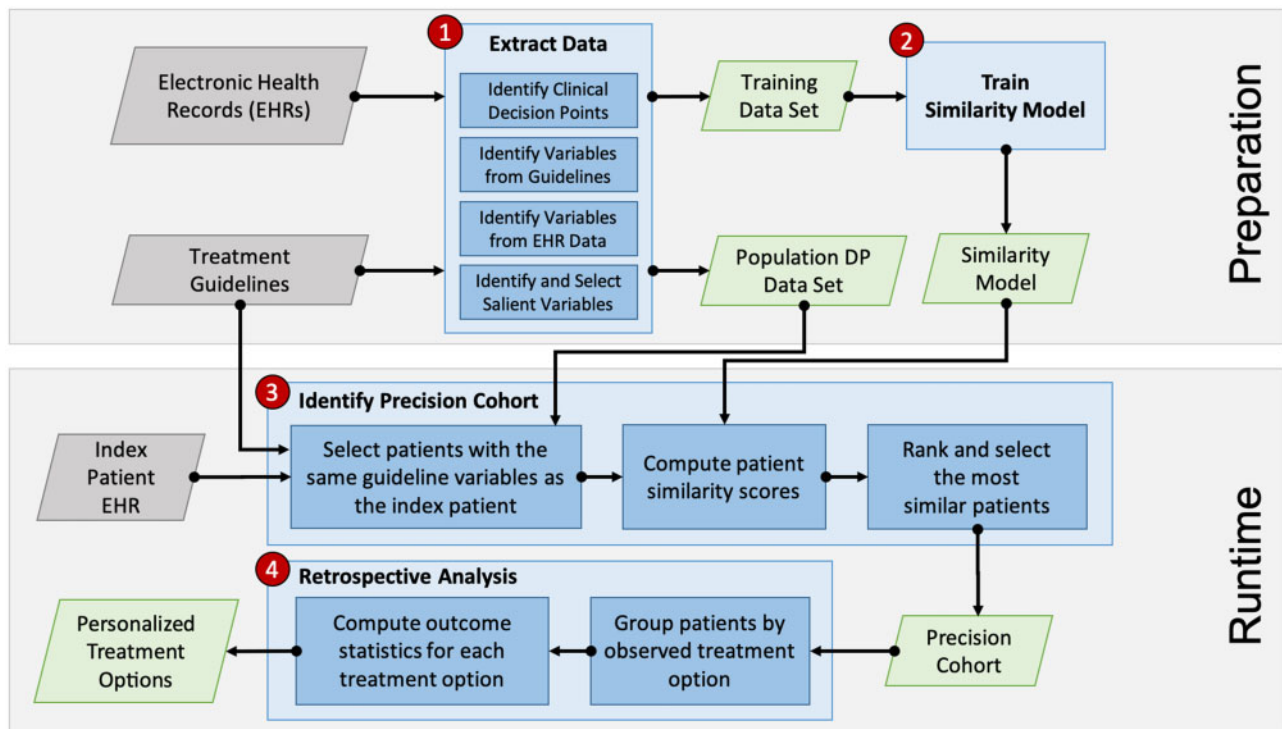
Figure 1. Overall Precision Cohort Treatment Options (PCTO) workflow consisting of 4 major steps: 1) data extraction, 2) similarity model training, 3) precision cohort identification, and 4) treatment options analysis. Steps 1 and 2 make up the preparation stage, which is performed offline and in advance. Steps 3 and 4 form the runtime stage, which is used in real time during the patient visit encounter.

Table 1. Decision point criteria and characteristics for the 3 disease conditions

	Hypertension (HTN)	Type 2 Diabetes Mellitus (T2DM)	Hyperlipidemia (HL)
Base treatment guideline	JNC7 and JNC-8 [9]	ADA [10]	AHA/ACC [11]
Study period (start - end) Rationale	1/1/2004–12/31/2018 Include data from JNC7 publication to present.	1/1/2008–12/31/2018	1/1/2008–12/31/2018
Decision point (DP) criteria	• Identify 2 consecutive uncontrolled BP readings (SBP $\geq$ 140 or DBP $\geq$ 90 mmHg) separated by 1–365 days. • Date of the second encounter is a decision point. • Age $\geq$ 18 at DP. • Not pregnant in prior 12 months.	• If the patient has a T2DM diagnosis on the problem list, the index date is the date of the diagnosis. If the patient does not have a T2DM problem list diagnosis, find the earliest 2 consecutive HbA1c $\geq$ 6.5% readings separated by 1–365 days, and the index date is the date of the second encounter. • All encounters after the index date with an associated HbA1c $\geq$ 7.0% within 365 days before the encounter date is a decision point. • Age $\geq$ 18 at DP. • Not pregnant in prior 12 months.	• Patient must have an HL diagnosis on the problem list. • All encounters after the HL diagnosis with an associated LDL > 130 mg/dl within 420 days before the encounter date is a decision point. • Age $\geq$ 18 at DP. • Not pregnant in prior 12 months.
Baseline period	Start of the baseline period depends on the specific variable. Some variables use data within the past 12 months, past 14 months, or all available history. The end of the baseline period is the decision point date.		
Baseline variables	The variable selection process described in Supplementary Section S2.1 was performed. The final set of selected variables for HTN, T2DM, and HL are described in Supplementary Tables S2–S4, respectively.		
Treatment options	List of clinically acceptable medication treatments was obtained via manual review of the relevant clinical treatment guidelines. The treatment variables for HTN, T2DM, and HL can be found in Supplementary Tables S2–S4, respectively.		
Follow-up period	1–365 (if no treatment change) or 14–365 (if treatment change) days after the decision point date	90–365 days after the decision point date.	30–450 days after the decision point date.
Outcome specification	First BP between N and 365 days after the decision point date. If no treatment changed, N = 0, otherwise N = 14. An SBP $\geq$ 140 or DBP $\geq$ 90 mmHg is considered not controlled.	First HbA1c lab test result between 90 and 365 days after the decision point date. An HbA1c $\geq$ 7% is considered not controlled.	First LDL lab test result between 30 and 450 days after the decision point date. An LDL > 130 mg/dl is considered not controlled.
Number of patients not in-control	157 942	24 373	63 510

Abbreviations: ACC, American College of Cardiology; ADA, American Diabetes Association; BP, blood pressure; DBP, diastolic blood pressure; DP, decision point; HbA1c, hemoglobin A1C; JNC, Joint National Committee; LDL, low-density lipid; SBP, systolic blood pressure.

The precision cohort

Figure 1 depicts the process used to create a "precision cohort." The approach was implemented as a data-analytic workflow consisting of 4 steps: 1) data extraction, 2) similarity model training, 3) precision cohort identification, and 4) treatment outcomes analysis. The unit of analysis is a decision point. At each touch point (eg, clinic encounter, telephone, or online encounter) where a patient's disease status was not controlled, the clinician had an opportunity to intervene. Refill encounters were excluded. We labeled each of those opportunities "decision points" (DPs). The DPs are drawn from all office encounters regardless of department and include primary care and specialists.

In the first step, the EHR data were processed and analyzed to extract clinical DPs. Each DP was associated with an index date, a preceding baseline period, and an observation period following the index date (Supplementary Figure S1). The selection criteria for the DPs, duration of the baseline and follow-up periods, specification of the baseline variables, identification of the treatment options, and definition of the outcomes are disease-specific and determined by clinical expert domain knowledge. The DP criteria and characteristics for the 3 disease conditions are described in Table 1. Data in the baseline period were used to extract baseline covariates that indicate health condition status (comorbidities, lab tests, procedures, etc)

and active treatments at the decision point. In addition, covariates associated with the propensity of receiving treatments based on clinical treatment guidelines for each disease were extracted. Actual treatment decisions administered at the decision point were identified. Data in the follow-up observation period were analyzed to determine the outcome associated with the treatment received (ie, disease controlled or uncontrolled). The control status of the disease is based on national guidelines as adopted by the medical group at the pilot site.[9–11] Although the hyperlipidemia guideline recommended a specific intensity of statin use rather than treating to a target LDL value, we made a conservative assumption that "moderate intensity" statin administration would reduce the highest untreated LDL treatment threshold in a patient without cardiac risk factors (190 mg/dl) by 30% (the guideline definition of moderate-intensity effect). Accordingly, we used the achievement of an LDL <130 mg/dl as the indication of being "controlled" (ie, undergoing statin treatment).

Detailed descriptions of the data extraction process are included in Supplementary Section S2.1. The result of the data extraction process is a data frame where the rows consist of individual decision points and the columns contain the treatment decision, the associated outcome, and the selected baseline variables. Only decision points with no missing variable values were retained. Next, the DPs

were partitioned to create a training data set and a population DP data set. The training data set was used to train a machine-learning model for computing patient similarity. The population DP data set included DPs from the entire patient population except for those used as part of the training data set. In an operational setting, the population DP data set can be updated frequently with new data captured in the EHR.

In the second step, a disease-specific similarity measure was derived from the training data set using a locally supervised metric learning algorithm that has been previously described[12] and applied to a large number of different use cases.[13–17] The similarity model not only selected highly influential covariates to significantly reduce the dimensionality but also learned covariate weights based on their relative effect on predicting patient outcomes. Detailed descriptions of the similarity model training process are given in Supplementary Section S2.2, and the final set of selected features and their corresponding similarity weights are listed in Supplementary Tables S2–S4 for HTN, T2DM, and HL, respectively.

In the third step, for the index patient, a precision cohort of clinically similar patient DPs from the population DP data set was dynamically created at the point-of-care. Causal inference matching methods, leveraging both exact match and similarity scoring, were used to adjust for baseline confounders.[18,19] We measured covariate balance to assess the likelihood of bias in the cohort used for the downstream treatment effect analyses between the "treatment change" and "no treatment change" groups.[20] Detailed descriptions of the precision cohort identification process are given in Supplementary Section S2.3.

In the final step, for each DP, a retrospective analysis was performed on the precision cohort to generate personalized treatment options relevant for the index patient. The analysis was done by first grouping the DPs by the actual treatment options selected at the time of the decision point and then for each treatment option group computing the size of the group, the associated outcome (percent controlled), the difference in the outcome compared to the "no treatment change" baseline option, and a statistical significance assessment of this outcome difference. Detailed descriptions of the treatment options analysis process are given in Supplementary Section S2.4.

We developed a visualization of the precision-cohort analysis that was presented to the practicing clinician in the EHR at the point-of-care. We began with an ethnographic study of clinicians' use of the EHR using a think-aloud protocol. After the initial design of the user interface was completed, we conducted 50 iterative feedback sessions on the evolving prototype. The final visualization appears as an integrated activity within the clinic's EHR and is shown in Figure 2.

Precision population analysis

The decision-point perspective can also be extended to populations. Traditional population-health reports represent the control status of patients at a slice-in-time. In reality, a patient's risk of complications accumulates during the total time the disease is not adequately controlled. Consequently, a patient's life journey with a chronic disease is better represented as a plot of disease control over time. Figure 3A depicts an example of the timeline of a single patient's disease status (eg, HbA1c for T2DM). Figure 3B–D presents summary views of a population's individual disease journeys over time plotted and stacked for T2DM, HL, and HTN, respectively. We refer to this fine-grain analysis of a population based on dynamically created precision cohorts as precision population analytics (PPA). The PPA report transforms the traditional population-health report of patients not-in-control at a single point in time into a graphical visualization of patient journeys over time. The visual impact of the dominant color portrays a better picture of the lives at-risk in a population; it highlights the amount of time not-in-control for each patient. It also raises awareness that there are multiple DPs in a patient's journey—something that is not well appreciated during any single encounter.

Figure 4 shows the overall PPA analysis workflow, which consists of 4 steps: 1) extracting the appropriate information from the EHR data to create DPs, 2) recording the actual treatment options (ATO) from the DPs, 3) computing the best-practice precision cohort treatment option (PCTO) for each decision point, and 4) analyzing the DPs, comparing the ATO versus PCTO and associated outcomes, and generating the PPA reports.

All available DPs were included in the population-level analysis. To accomplish this, a standard leave-one-out cross-validation (LOOCV) approach[21] was adopted. Each decision point in the population DP data set was selected for removal to create an updated population DP data set. For each DP, the ATO followed at each decision point is recorded along with the associated outcome (controlled or not controlled). Then, using the updated population DP data set, all possible PCTOs at each decision point are computed along with their associated estimated outcomes (percent controlled) using the precision cohort workflow (Figure 1, steps 3 and 4) described previously. A single PCTO is then selected by identifying the statistically significant best treatment option, if available. If there are no treatment options that resulted in statistically significantly better control, the no-change option (stay the current course) is the default case, and there is no PCTO for that DP. Supplementary Tables S8–S10 list the most frequent PCTOs generated (along with their associated average percent controlled) for HTN, T2DM, and HL, respectively.

A sensitivity analysis was performed by removing all DPs that belong to the same patient as the selected DP in the LOOCV processing (instead of removing just the single selected DP). No significant change in the overall results was observed. There are several possible reasons for this: 1) since the DPs are from different points in time in the patient's disease trajectory, they are not very similar to each other and are rarely included in the final precision cohort; 2) since a typical precision cohort consists of thousands of DPs, the exclusion of a few DPs will likely have a small impact.

The outcomes of the ATO and PCTO are compared to quantify differences in expected vs actual control. The PPA analysis can be performed for the entire medical-group patient population or for a specific cohort, such as an individual physician's panel of patients.

All of the computational workflows and analyses were implemented using a combination of SQL, Python, and R.[22,23]

RESULTS

The PPA analysis workflow described above was run on the deidentified EHR data and the results summarized in Table 2 for each disease (HTN, T2DM, HL). For the majority of DPs, there was a treatment option that led to statistically better control results for similar patients in similar situations than the actual chosen treatment option, which was the case for 66.8%, 59.0%, and 83.5% of DPs for HTN, T2DM, and HL, respectively. The expected outcome if the organization's best-practice PCTO had been chosen at each DP would have led to improved outcomes over the ATO: 65.1% vs 48.0% (HTN), 37.7% vs 22.5% (T2DM), and 75.3% vs 31.7%
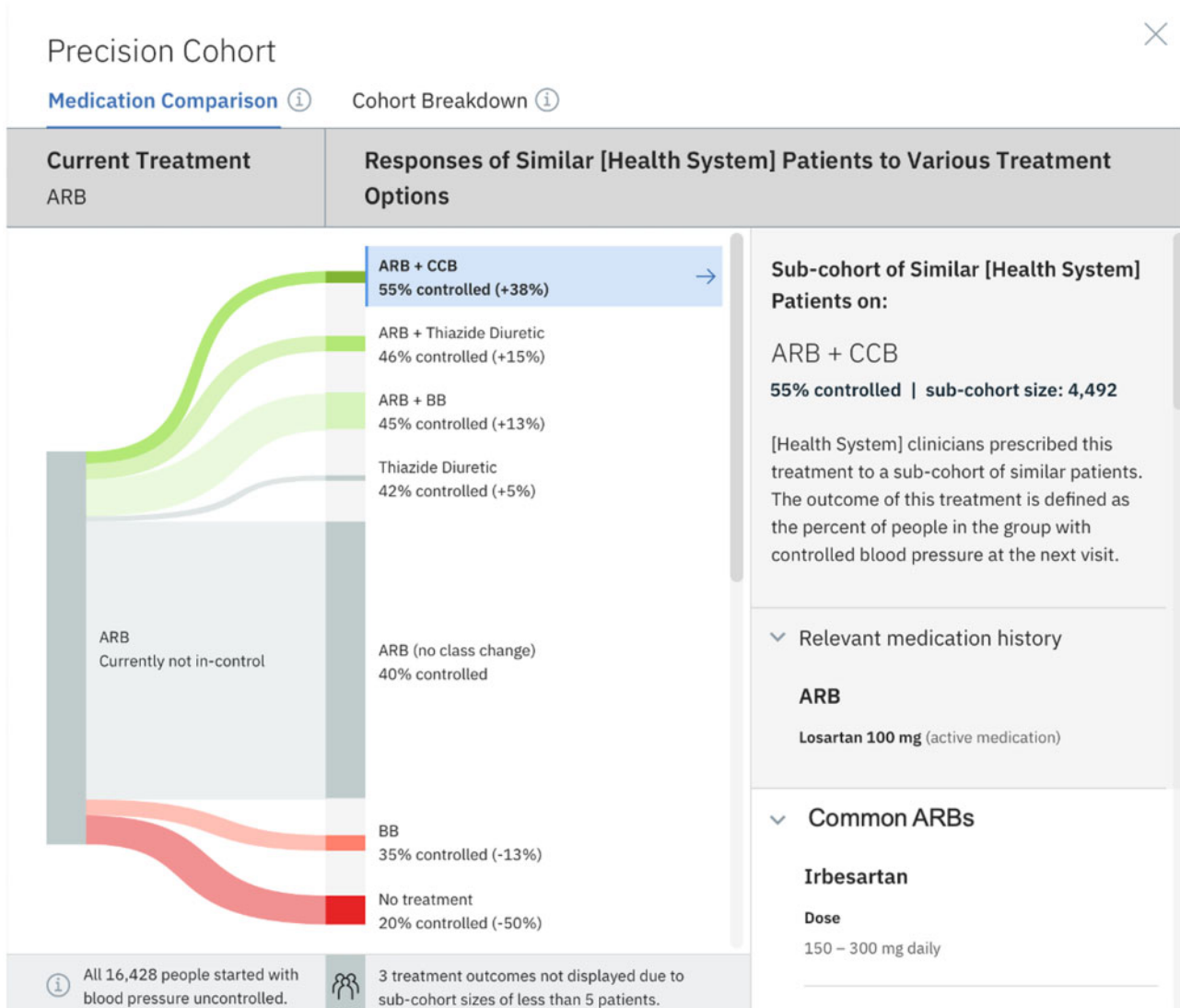
Figure 2. Precision cohort visualization in the EHR. This diagram depicts observed outcomes for a precision cohort of patients who are similar to the individual patient under the same clinical situation (defined by the similarity model). In this example, all patients in the cohort have a diagnosis of hypertension and are treated with an angiotensin receptor blocker (ARB) only (column 1). In the middle column, other clinicians in the practice have chosen multiple treatment options. The width of the "prong" shows the relative size of the cohort choosing that option. The most common choice was to make no change in the ARB drug class. Under the no-change scenario, only 40% of those patients had controlled blood pressure (BP) at the follow-up measurement. The prongs above the no-change group all had an increase in percent controlled on follow-up. The treatment cohorts in green had a statistically significant change in percent controlled. The prongs below the no-change group had a lower percent controlled on follow-up, and red prongs indicate a statistically significant change.

(HL). This would represent a hypothetical improvement in percent controlled of 36%, 68%, and 138%, respectively.

DISCUSSION

Sackett et al defined evidence-based medicine as the "conscientious, explicit, and judicious use of current best evidence… integrating individual clinical expertise with the best available external clinical evidence from systematic research." He went on to say, "Good doctors use both individual clinical expertise and the best available external evidence, and neither alone is enough."[24] Up until 10 years ago, before the enactment of the HITECH (Health Information Technology for Economic and Clinical Health) legislation, results of clinical trials were the primary evidence available to formulate clinical guidelines to assist with medical decision-making, given the low

penetrance of EHRs at the time. With the availability of comprehensive clinical data in EHRs, machine-learning analytical tools, and relevant experiences of past patients in similar situations, precision cohorts can be presented to better inform clinical decision-making, potentially improving the percentage of patients whose diseases are under control by 36%–138% for the conditions we studied. These real-world best practices complement evidence produced through RCTs, bridging the gap between efficacy reported in the controlled environment of a clinical trial and the actual effectiveness observed in the real world.

One of the barriers to change is clinical inertia, which is a complex interaction amongst the patient, clinician, and health system environment.[25] In our study, the most common action at a DP was no change. Part of behavioral change begins with raising the awareness of the problem for all parties and providing accessible means for
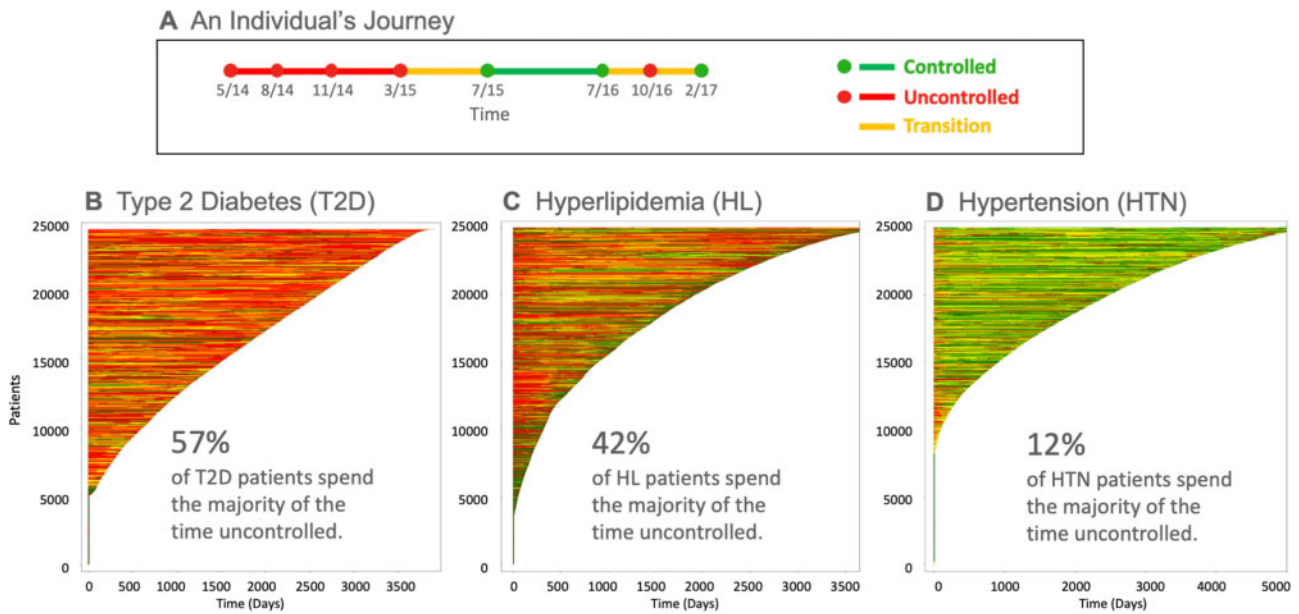
Figure 3. Population of patient journeys for 3 chronic conditions. A) An individual patient's journey. Green dots represent encounters at which the disease parameter (eg, HbA1c) is controlled. Red dots represent encounters at which the disease parameter is not controlled. Red lines connect 2 consecutive encounters with uncontrolled outcomes. Green lines connect 2 consecutive encounters with controlled outcomes. Yellow lines connect consecutive encounters with different outcomes (1 controlled and 1 uncontrolled). B, C, D) Journeys for 25 000 randomly selected type 2 diabetes, hyperlipidemia, and hypertension patients, respectively. The individual patient journeys are stacked vertically and sorted in descending order by the duration of the patient's longitudinal observations (days).
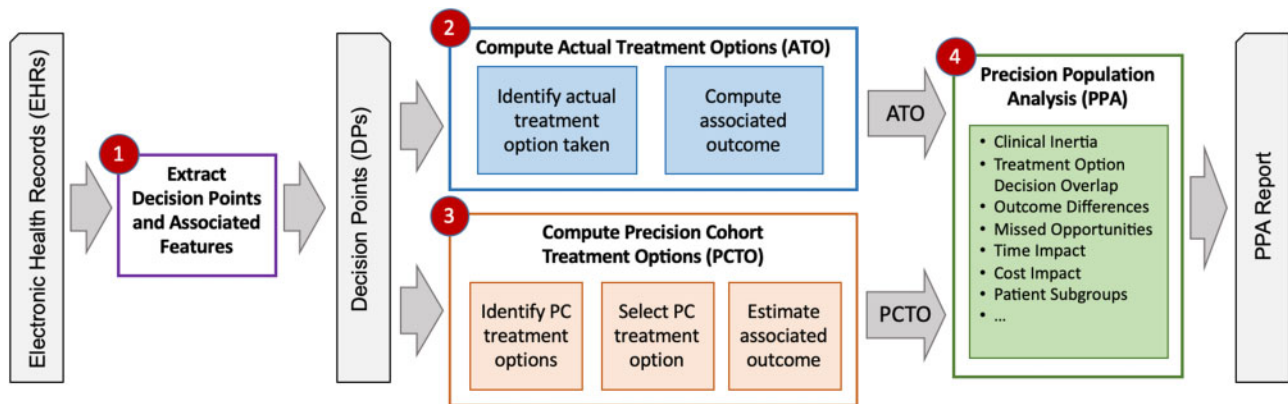
Figure 4. Overall precision population analytics (PPA) workflow consisting of 4 steps: 1) extracting decision points (DPs) and associated features from the EHR data; 2) computing the actual treatment options (ATOs) from the DPs, 3) computing the precision cohort treatment options (PCTO) from the DPs, and 4) analyzing the DPs, comparing the actual versus precision cohort treatment options and associated outcomes, and generating the PPA reports.

changing.[26] A PPA analysis of an individual physician's panel provides a graphic visualization of the panel's cumulative risk over time, which can either pinpoint opportunity areas for intervention or highlight exemplary achievement. In either case, specific comparison of personal performance to local organizational cohorts serves to motivate behavioral change for both the patient and the clinician.

Longhurst et al[27] proposed a "green button" functionality in EHRs to provide access to practice-based evidence derived from the analysis of aggregate data within the EHR. Our methods and application are consistent with the vision they described. We have developed and deployed our application within an EHR system integrated within the physician's workflow. The precision-cohort visualization is available at the point-of-care tailored to each patient and his or her specific disease state.

There are limitations to the use of observational data. Without careful attention to data quality and methodological integrity, machine-learning models could be tainted by bias. As described in the Methods section, the precision-cohort workflow leverages several approaches to identify appropriate baseline confounders including covariates capturing the health condition and active treatment status at the decision point, covariates associated with the propensity of receiving treatment based on clinical guidelines, and covariates associated with the outcome of interest selected using data-driven methods from the EHR data. Causal inference matching methods, leveraging both exact match and similarity scoring, are then used to adjust for baseline confounders, and covariate balance is assessed to ensure that the precision cohort is not biased and is appropriate for treatment-effect analyses. Even with these mitigation

Table 2. Precision Population Analytics (PPA) results for the 3 disease conditions: hypertension, type 2 diabetes mellitus, and hyperlipidemia

	HTN	T2DM	HL
Number of patients	157 942	24 373	63 510
Number of decision points (DPs)	733 300	171 203	561 971
Percent of DPs with uncontrolled outcome	51.8%	81.2%	64.1%
Percent of DPs with a significant PCTO	69.1%	60.3%	84.4%
Percent of DPs with a significant PCTO that is not the ATO	66.8%	59.0%	83.5%
Avg (SD) outcome (percent controlled) for these DPs when following the ATO	48.0% (50.0)	22.5% (42.0)	31.7% (47.0)
Avg (SD) outcome (percent controlled) for these DPs when following the PCTO	65.1% (7.8)	37.7% (12.2)	75.3% (4.7)

Abbreviations: ATO, actual treatment option; Avg, average; DP, decision points; HL, hyperlipidemia; HTN, hypertension; PCTO, precision cohort treatment option; SD, standard deviation; T2DM, type 2 diabetes mellitus.

methods, the selection of confounder covariates may not be fully complete, and biases may not be completely eliminated.

In the current work, only medication-based treatment options were considered. Since nothing in the method explicitly precludes handling additional treatment types, a natural extension is to include nonpharmacologic treatment options such as diet, exercise, smoking cessation, and other lifestyle modifications or interventions to improve unfavorable social determinants of health as long as the necessary data are available. This study only used data from a single health system, and the decision points were extracted from historical outpatient data which may potentially contain outdated treatment information and historical biases.[28] We attempted to mitigate the impact of this by adjusting the lookback period to coincide with the publication date of the most recent national guidelines and only considered drugs approved by current guidelines.

Decision points were explicitly defined as an opportunity to act on a disease not under control. Sometimes the focus of a specific visit may not include 1 of the 3 conditions we studied. From a patient health risk point-of-view, however, a deferred action does affect the patient's cumulative health risk, so we felt it was accurate to consider each encounter with 1 of the conditions uncontrolled as a decision point regardless of whether it was the focus of the visit.

Although using data from approximately 2.5 million community patients from a single large health system was sufficient to provide statistically significant treatment options for the common diseases we studied, there will be more limited numbers of similar patients for less common diseases. One way to increase the total number of patients in the pool is to combine data from multiple health systems. Applying the methods, which we developed to data from additional health systems, is needed to assess the generalizability of this approach. Combining data from different sites, however, introduces other confounders that may decrease the precision of the similarity of patients within a given cohort. For example, the fact that the precision cohorts were created from data of patients in the same catchment area with treatment options chosen by local professional colleagues appeared to influence the perception by clinicians of the applicability of the precision cohort analysis to their decisions. This

tradeoff should be explored in future studies to assess the importance of using local data on the behavior of local clinician decision makers.

With the advent of widespread use of EHRs and many other electronic sources of health-related data, physicians are faced with voluminous sources of data which contribute to physician burnout.[29] Instead of inundating physicians with overwhelming amounts of raw data, analytical tools should be provided to help extract and visualize the insights contained in those data. Providing physicians with a summary analysis about the best-practice experiences of their colleagues could help them make more informed decisions—a continuously learning health system[30] brought to the point-of-care.

CONCLUSION

Clinicians seek to recommend the best treatment decisions for each individual patient. Up until now, evidence-based medicine primarily relied on the results of RCTs which, by their nature, focus on homogeneous populations. Complementing clinical guidelines with insights from real-world data of precision cohorts of similar patients offer clinicians additional data to tailor treatment decisions for individual patients with the potential to significantly improve the control of common chronic diseases.

FUNDING

AUTHOR CONTRIBUTIONS

All coauthors contributed to the design and implementation of the pilot, participated in the manuscript preparation, and approve of and agree to be accountable for the accuracy and integrity of the publication.

SUPPLEMENTARY MATERIAL

Supplementary material is available at *Journal of the American Medical Informatics Association* online.

ACKNOWLEDGMENTS

CONFLICT OF INTEREST STATEMENT

The coauthors, except for John Zambrano, were employees of IBM during the time the research was conducted. Paul Tang's current affiliation is at Stanford University.

REFERENCES

1. Kent DM, Hayward RA. Limitations of applying summary results of clinical trials to individual patients: the need for risk stratification. *JAMA* 2007; 298 (10): 1209–12.
2. Eddy DM, Adler J, Patterson B, *et al*. Individualized guidelines: the potential for increasing quality and reducing costs. *Ann Intern Med* 2011; 154 (9): 627.

3. Kennedy-Martin T, Curtis S, Faries D, *et al*. A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results. *Trials* 2015; 16 (1): 495.

4. Wright JT, Williamson JD, Colla CH, *et al*. A randomized trial of intensive versus standard blood-pressure control. *N Engl J Med* 2015; 373: 2103–16.

5. Unger JM, Cook E, Tai E, *et al*. Role of clinical trial participation in cancer research: barriers, evidence, and strategies. *Am Soc Clin Oncol Educ Book* 2016; 36: 185–98.

6. Basu S, Sussman JB, Hayward RA. Detecting heterogeneous treatment effects to guide personalized blood pressure treatment: a modeling study of randomized clinical trials. *Ann Intern Med* 2017; 166 (5): 354–60.

7. Cantor MN, Thorpe L. Integrating data on social determinants of health into electronic health records. *Health Aff (Millwood)* 2018; 37 (4): 585–90.

8. Walker RJ, Strom Williams J, Egede LE. Influence of race, ethnicity and social determinants of health on diabetes outcomes. *Am J Med Sci* 2016; 351 (4): 366–73.

9. James PA, Oparil S, Carter BL, *et al*. 2014 evidence-based guideline for the management of high blood pressure in adults: report from the panel members appointed to the Eighth Joint National Committee (JNC 8). *JAMA* 2014; 311 (5): 507–20.

10. American Diabetes Association. Pharmacologic approaches to glycemic treatment. *Diabetes Care* 2017; 40: S64–74.

11. Stone NJ, Robinson JG, Lichtenstein AH, et al. 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol* 2014; 63: 2889–934. doi:10.1016/j.jacc.2013.11.002

12. Wang F, Sun J, Li T, *et al*. Two heads better than one: metric + active learning and its applications for IT service classification. In: Ninth IEEE International Conference on Data Mining, 2009. ICDM '09. 2009. 1022–7. doi: 10.1109/ICDM.2009.103

13. Wang F, Zhang C. Feature extraction by maximizing the average neighborhood margin. In: IEEE Conference on Computer Vision and Pattern Recognition, 2007. CVPR '07. 2007. 1–8. doi: 10.1109/CVPR.2007.383124

14. Sun J, Wang F, Hu J, *et al*. Supervised patient similarity measure of heterogeneous patient records. *SIGKDD Explor Newsl* 2012; 14 (1): 16–24.

15. Liu H, Li X, Xie G, *et al*. Precision cohort finding with outcome-driven similarity analytics: a case study of patients with atrial fibrillation. *Stud Health Technol Inform* 2017; 245: 491–5.

16. Ng K, Sun J, Hu J, *et al*. Personalized predictive modeling and risk factor identification using patient similarity. *AMIA Jt Summits Transl Sci Proc* 2015; 2015: 132–6.

17. Wang F, Sun J. PSF: a unified patient similarity evaluation framework through metric learning with weak supervision. *IEEE J Biomed Health Inform* 2015; 19 (3): 1053–60.

18. Stuart EA. Matching methods for causal inference: a review and a look forward. *Stat Sci* 2010; 25 (1): 1–21.

19. King G, Nielsen R. Why propensity scores should not be used for matching. *Polit Anal* 2019; 27: 435–54. doi:10.1017/pan.2019.11

20. Sauppe JJ, Jacobson SH. The role of covariate balance in observational studies. *Naval Res Logist* 2017; 64 (4): 323–44.

21. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection In: *Proceedings of the 14th International Joint Conference on Artificial intelligence—Volume 2*. San Francisco, CA: Morgan Kaufmann; 1995: 1137–43. http://dl.acm.org/citation.cfm?id=1643031.1643047 Accessed July 12, 2013

22. Pedregosa F, Varoquaux G, Gramfort A, *et al*. Scikit-learn: machine learning in Python. *J Mach Learn Res* 2011; 12: 2825–30.

23. R Development Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2008.

24. Sackett DL, Rosenberg WM, Gray JA, *et al*. Evidence based medicine: what it is and what it isn't. *BMJ* 1996; 312 (7023): 71–2.

25. Okemah J, Peng J, Quiñones M. Addressing clinical inertia in type 2 diabetes mellitus: a review. *Adv Ther* 2018; 35 (11): 1735–45.

26. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. *Am J Health Promot* 1997; 12 (1): 38–48.

27. Longhurst CA, Harrington RA, Shah NH. A "green button" for using aggregate patient data at the point of care. *Health Aff* 2014; 33 (7): 1229–35.

28. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 2019; 366 (6464): 447–53.

29. Melnick ER, Dyrbye LN, Sinsky CA, *et al*. The association between perceived electronic health record usability and professional burnout among US physicians. *Mayo Clin Proc* 2019. doi: 10.1016/j.mayocp.2019.09.024.

30. Institute of Medicine. *Best Care at Lower Cost: The Path to Continuously Learning Health Care in America*. Washington, DC: The National Academies Press; 2013.