NEJM Catalyst | Innovations in Care Delivery

ARTICLE

# Using Aggregate Patient Data at the Bedside via an On-Demand Consultation Service

Alison Callahan, PhD, MIS, Saurabh Gombar, MD, PhD, Eli M. Cahan, Kenneth Jung, PhD, Ethan Steinberg, Vladimir Polony, Keith Morse, MD, MBA, Robert Tibshirani, PhD, MS, Trevor Hastie, PhD, MS, Robert Harrington, MD, Nigam H. Shah, PhD, MBBS

Vol. 2 No. 10 | October 2021

DOI: 10.1056/CAT.21.0224

Using evidence derived from previously collected medical records to guide patient care has been a long-standing vision of clinicians and informaticians and one with the potential to transform medical practice. As a result of advances in technical infrastructure, statistical analysis methods, and the availability of patient data at scale, an implementation of this vision is now possible. Researchers at Stanford University developed an on-demand consultation service to derive evidence from patient data to answer clinicians' questions and support their bedside decision-making. The authors describe the design and implementation of the service as well as a summary of their experience in responding to the first 100 requests. Consultation results informed individual patient care, resulted in changes to institutional practices, and motivated further clinical research. Stanford makes the tools and methods publicly available to facilitate the broad adoption of such services by health systems and academic medical centers.

Evidence-based medicine emphasizes the "conscientious, explicit, and judicious use of current best evidence"[1] when making treatment decisions.[2,3] Randomized controlled trials (RCTs) are considered the highest-quality source of evidence about treatment efficacy and safety. Evidence derived from RCTs, however, often does not generalize to the majority of patients, who tend to have multiple comorbidities, take many medications, and differ from individuals enrolled in RCTs on many characteristics,[4] resulting in an inferential gap between the evidence that is available and that which is needed.[5,6] Therefore, it is necessary to transform the

evidence-generation process[7] and to incorporate the use of aggregate patient data at the point of care[8] in order to create a successful learning health system.[9]

Electronic medical records (EMRs) are a source of rich longitudinal data about millions of real-world patients. Since the 1970s,[10,11] clinicians and scientists have envisioned using the medical records of previously treated patients to inform the care of current and future patients. A more recent example from 2011 described the use of EMR data to support the management of an adolescent female with systemic lupus erythematosus; at the time, however, incorporating data from EMRs into clinical decision-making required significant manual effort to verify context and applicability, rendering it infeasible for use in routine patient care.[12]

Today — with the adoption of EMRs across the United States and internationally, the increasing ease of use of advanced statistical methods, and the ability to compute with large patient cohorts — a core tenet of the learning health system has been enabled: deriving on-demand evidence for diverse clinical scenarios from the EMR.[7,13]

Using these advances as a foundation, the authors designed, developed, and offered a consultation service that used EMR and medical insurance claims data at Stanford Medicine to provide on-demand evidence for questions arising during clinical care.[14] In this article, the authors report their findings from responding to the first 100 requests to the service: they summarize requests by medical specialty, the types of analyses required to fulfill their requests, and clinicians' responses to the evidence returned. Their service responded to requests from 53 users from multiple specialties; 83 consultations were completed, of which 48 were descriptive analyses and 35 were treatment comparison analyses. Ten consultations led to changes to patient care, 52 guided further research, and 17 led to follow-up analyses, including four that were presented at medical conferences or published in peer-reviewed journals.

## The Setup of the Consultation Service

Beginning in February 2017, with approval from the Stanford Institutional Review Board, we offered a consultation service to provide on-demand evidence to clinicians at Stanford Medicine, staffed by a team of four (a clinical informatics–trained fellow, two EMR data specialists, and a part-time data scientist). The service availability was announced via grand rounds in the departments of Medicine, Pediatrics, and Dermatology and in presentations to trainees and faculty in the Urology, Pathology, and Neurology departments, as well as to the Chief Residents Council. After this initial outreach to residents, fellows, and attending physicians, knowledge of the service spread via word of mouth. As part of offering the service, we collected data on the motivations for consultation requests and on the subsequent actions taken in light of the evidence returned. On reaching 100 requests (in September 2018), we analyzed the consultation request motivations and resulting actions, assessed the concordance of consultation results across clinical data sources, and quantified the expected false-positive rate. These analyses concluded in August 2019.

In designing the service, we leveraged best practices,[15] methods,[16] and tools[17,18] to derive evidence from EMRs. We have summarized recommendations for conducting and reporting

observational studies performed using EMRs derived from a large body of our team's prior work.[15] For example, we have used historical EMR data for vigilance, such as monitoring adverse drug events[19,20] and surveilling implantable devices[21]; for answering clinical questions, such as whether there is an association between androgen deprivation therapy and dementia;[22,23] and for elucidating quality of care, by profiling unplanned ED visits,[24] surfacing patient reported outcomes,[25] and quantifying treatment variability in metastatic breast cancer.[26] We have also learned from leading collaborative studies,[27] developing methods for electronic phenotyping,[28–31] and participating in multiple Observational Health Data Sciences and Informatics network studies.[32–38]

> *"The consultation service setup is designed to receive questions from clinicians, retrieve the appropriate patient data using a specialized search engine, perform the analyses required for the question, and return a report summarizing the results.*

The consultation service setup is designed to receive questions from clinicians,[14] retrieve the appropriate patient data using a specialized search engine,[18] perform the analyses required for the question, and return a report summarizing the results. Methods for data extraction, processing, and analysis are used in the consultation service.[16] A platform for clinical data science at Stanford Medicine supported the operation of the service.[17]

## The Workflow for Fulfilling a Consultation Request

A consultation began with an email from a requestor, detailing a clinical question. Upon receiving the request, the informatics clinician member of our dedicated four-person team scheduled an intake discussion with the requesting clinician to specify the population, intervention, comparator, outcome, and timeframe (PICOT) for their particular question.[14]

On the basis of the PICOT formulation of the question, the team's EMR data specialist constructed patient cohorts using the Advanced Cohort Engine (ACE)[18] to search one of three data sources: EMRs of 3.1 million individuals from Stanford Medicine; IBM MarketScan insurance claims for 124 million individuals; or Optum Clinformatics Data Mart insurance claims for 53 million individuals. The data scientist then conducted the necessary statistical analyses and worked with the informatics clinician to write a report summarizing the analyses and their results. The report was then shared with the requestor and explained during an in-person, structured debrief session. Each report consisted of the original question as posed, the PICOT reformulation, and sections summarizing the cohort demographics and the interpretation of the analyses, with a detailed walkthrough of the analyses. Three example reports and the debrief script are provided in the Appendix. The interaction was designed to be similar to obtaining a second opinion from a colleague.

Our workflow evolved to incorporate real-time searches of the EMR as the informatics clinician collected PICOT details. For example, if a given cohort criterion returned very few patients,

then the informatics clinician could relay this information during the intake interview in order to elicit modifications to the cohort definition from the requestor. Clarifications needed during debrief interviews were also incorporated into subsequent reports and debriefs to better contextualize analysis results for requestors. The majority of this evolution occurred during the first 3 months of offering the service (Figure 1).

Over the course of the 18-month pilot study of the first 100 consultations, the median consultation turnaround time was 5 days. As the team gained experience and the service workflow matured, by the end of the study, reports were returned within 48 hours. The average time devoted to each report ranged from 3.75 hours to 5.75 hours. On the basis of this information, we believe a team composed of one full-time clinical informatics fellow, two full-time EMR data specialists, and a 20% part-time data scientist would be able to complete up to 20 such consultations in 1 week. The personnel costs for our geographic area (San Francisco Bay area of California) for this team are estimated at $505,000 per year. Yearly data access infrastructure, cloud compute, licensing, and professional service expenses come to an additional $70,000 per year. With these assumptions, the cost of running such a service would be approximately $550 per consultation.

## Data Sets and Cohort Building

The service used demographics, diagnoses, procedures, medications, laboratory values, clinical notes, length of stay, and mortality information for millions of patients from three data sources: EMRs from 3.1 million Stanford Medicine (Stanford) patients (54% female, spanning 1995–2019)[39] consisting of diagnosis, procedure, medication, and laboratory test records, as well as clinical notes processed using a previously developed and evaluated text-processing pipeline;[40,41] IBM MarketScan (MarketScan), which contains employer and Medicare insurance claims for 124 million lives (53% female, spanning 2007–2015); and Optum Clinformatics Data Mart (Optum), which contains insurance claims for 53 million lives from employer-sponsored health plans (53% female, spanning 2003–2016).

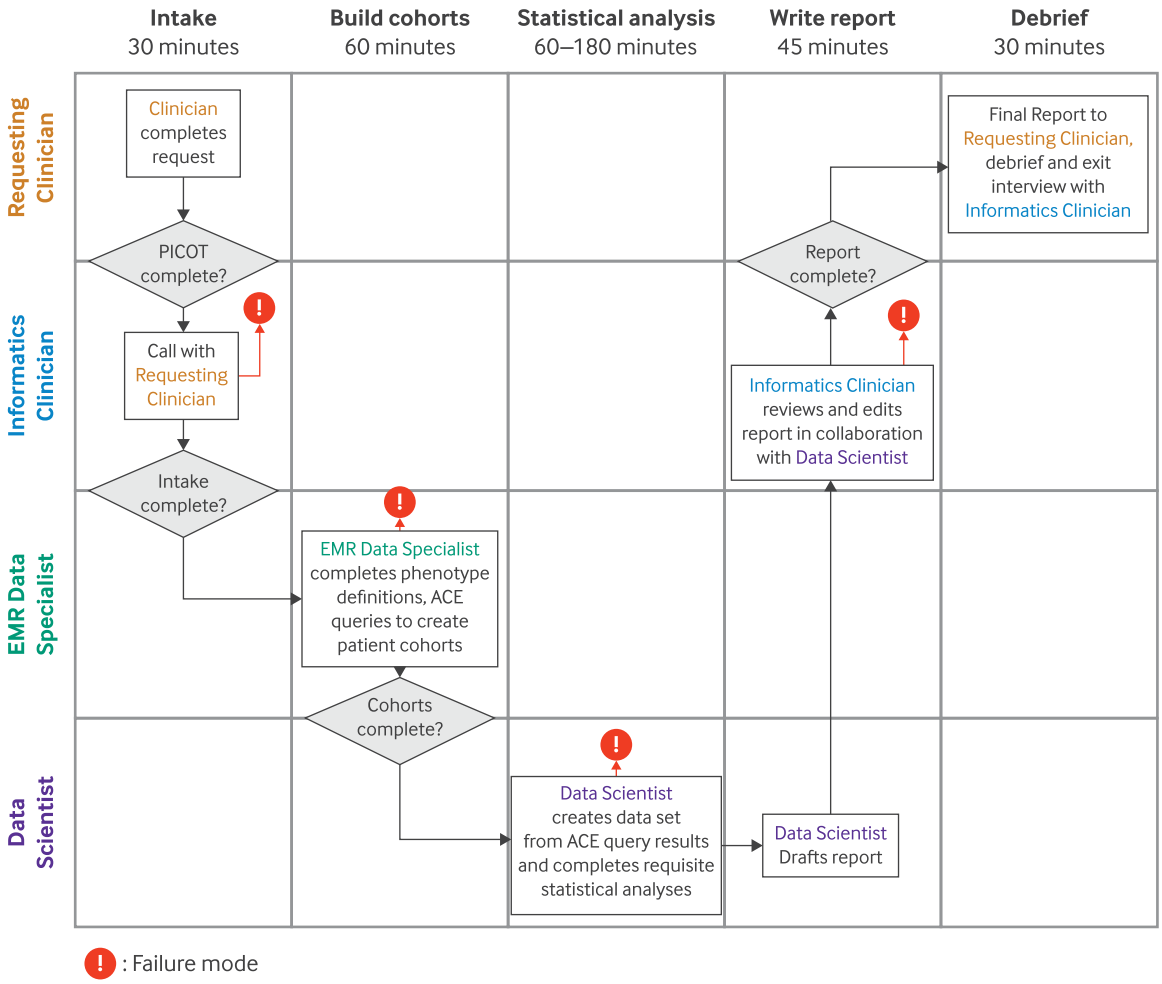
> *“The interaction was designed to be similar to obtaining a second opinion from a colleague.*

The choice of data set for a given consultation was informed by the question and was primarily based on meeting the criteria specified in the PICOT. For example, if a patient cohort definition relied on a specific range of laboratory test result values, then this necessitated using the Stanford EMR data set, because claims data do not include laboratory test results. The EMR data specialist constructed patient cohorts using the ACE[18] to define necessary and sufficient conditions to determine if an exposure or outcome of interest occurred in a patient’s timeline. A patient timeline view of patient records provided by ACE enabled anonymized chart review for quality checks of the exposure and outcome definitions and resulting cohorts.

FIGURE 1.

## Workflow for Fulfilling a Consultation Request

This figure illustrates the process of fulfilling each step in a consultation request, including the time required and how the team members interact with each other and the requesting clinician. On average, the time spent per request ranged from 3.75 to 5.75 hours. The team handled more than one request at a time, and the average turnaround time was less than 48 hours.




Note: Failure mode indicates that the process could not continue for some reason; in such cases, the team would provide a report to the requesting clinician containing the information gathered up to that point and explaining the reason that the team was unable to proceed further.

Abbreviations: ACE = Advanced Cohort Engine, EMR = electronic medical record, PICOT = population, intervention, comparator, outcome, and timeframe.

Source: The authors

## Supported Analyses

The service supported treatment comparisons for discrete, continuous, and time-to-event outcomes as well as for custom descriptive analyses.[16] For discrete, continuous, and time-to-event outcomes, we used a standardized process that attempted to emulate a target trial[42] on the basis of the criteria specified in the PICOT. For consultations requesting treatment comparisons, we created cohorts of similar patients using two approaches: Mahalanobis distance with a fixed caliper based on age, sex, length of record, and year of entry into the cohort (we call this simple matching) or high-dimensional propensity score matching (hd-PSM).[43,44] Matching is a way to identify subsets of patients who are similar in most respects other than in the treatment they received, in order to reduce the chance that observed differences in outcomes are caused by variation in properties other than treatment but that also impact the outcome (commonly referred to as confounding).[43] For propensity score estimation, we used an L2 regularized logistic regression model with a time-binned count–based featurization of pretreatment clinical data elements (diagnoses, procedures, and medication records), fit using glmnet.[45] Regularization strength was determined using 10-fold cross validation with a final refit on the entire data before estimating propensity scores for all patients.[46] Results from both matching strategies were included with each report.

The subsequent analysis performed on matched cohorts was selected on the basis of the outcome specified in the PICOT formulation of the question. For treatment comparisons with binary outcomes, we calculated odds ratios and associated confidence intervals. For treatment comparisons with continuous outcomes, we fit regression models and reported mean change in response estimates and associated confidence intervals. For treatment comparisons with time-to-event (survival) outcomes, we computed Kaplan-Meier plots and performed log-rank tests for differences in survival curves between compared treatments and reported hazard ratios (HRs) and associated confidence intervals.

Custom descriptive analyses required bespoke code for each request, primarily written in R, with data aggregation using Python as necessary. All analyses were conducted in R and are available for noncommercial use upon request.

## Quality Checks for Supported Treatment Comparison Analyses

Given that there is no known ground truth for the questions received by the consultation service, we established code correctness using synthetic data sets and derived an estimate of the false-positive rate for treatment comparison analyses using publicly available data sets of drug-effect pairs as ground truth.

### *Establishing Code Correctness*

We generated eight synthetic data sets, each with 10,000 patients, using all combinatorial variations of three properties: whether a binary treatment had an effect on a single survival outcome, whether treatment assignment had a dependence on a single binary covariate, and whether the covariate had an effect on the survival outcome. We confirmed the correctness of the analysis code by verifying that the analyses returned a treatment effect if and only if the underlying

data were constructed with a treatment effect and that the direction of the derived treatment effect was concordant with the treatment effect specified when creating the synthetic data set. On performing treatment comparisons using cohorts matched with hd-PSM, the analysis code correctly identified protective effects for the four synthetic data sets constructed to have such intervention effects and no effects for the four synthetic data sets constructed to have no effect. For the two synthetic data sets in which there was both a biased treatment and a covariate effect, resulting in confounding, propensity matching correctly recovered the true effect for the treatment.

### *Quantifying the Expected False-Positive Rate*

We selected treatment-outcome pairs known to be either associated or nonassociated as compiled by the Observational Medical Outcomes Partnership (OMOP) community[47] and the Exploring and Understanding Adverse Drug Reactions (EU-ADR) Project[48] because they are publicly available and have been used as ground truth sets in other studies.[49–51] While both reference sets contain known associations as well as asserted nonassociations, only the asserted nonassociations were informative in quantifying the false-positive rate. Of the 399 drug-outcome pairs from the OMOP community reference set, 234 are nonassociations. Of the 93 pairs from the EU-ADR project, 50 are nonassociations.

> *Over the course of the 18-month pilot study of the first 100 consultations, median consultation turnaround time was 5 days. As the team gained experience and the service workflow matured, by the end of the study, reports were returned within 48 hours.*

We constructed cohorts corresponding to each of the asserted nonassociated treatment-outcome pairs in the reference sets and estimated a treatment effect, using Stanford EMR data. Cohorts were constructed by transforming each treatment and outcome definition into corresponding ACE queries. Outcomes were defined using *International Classification of Diseases*, 9th Edition, Clinical Modification codes, and drug treatments were defined using RxNorm codes. We used a new patient cohort design in which patients were entered into a cohort immediately after the first time they were prescribed a drug. Outcomes were measured as events after the first prescription, with patients being marked as censored when their medical records ended. A result was counted as false-positive if our analysis found that a given treatment was associated with an increased or decreased HR relative to the comparator (with an effect estimate greater than or less than 1 and a $P$ value $\leq$.05), and the reference set marked it as not associated.

Of the 234 nonassociated pairs from the OMOP community, there were 137 drug-outcome pairs for which a minimum 100 patients exposed to the drug were present in Stanford data. Of these, 27 associations were false-positives and the remaining 110 were correctly identified as nonassociations, providing an estimated false-positive rate of 20%. From the 50 nonassociated treatment-outcome pairs from the EU-ADR project, there were 42 pairs for which there were enough data. Of these, seven associations were false-positives and the remaining 35 correctly identified as nonassociations, providing an estimated false-positive rate of 17%.

Because the OMOP and EU-ADR reference sets were constructed to evaluate methods for treatment comparisons, the 17%–20% expected false-positive rate is applicable to consultations requesting a comparison of the HR of an outcome between treatments.

## Summarizing the First 100 Consultation Requests

### *Categorizing Motivations for Requests and Subsequent Actions*

We categorized the scenarios motivating consultation requests and subsequent actions by requestors on the basis of the intake and debrief meetings, respectively. Each consultation request was assigned a single motivation category and one or more subsequent action categories. We categorized subsequent actions into one or more of three possible categories. If, during debrief, the requestor stated that they would use the knowledge gained from the consultation to change the treatment of a current or future patient with similar presentation, the consultation was categorized as having changed patient care. Debriefs in which the requestor planned to obtain approval to further study their question or to use the findings from the consultation to generate hypotheses for an ongoing research project were categorized as guiding further research. Debriefs in which the requester used the results from the consultation report directly as the basis of a publication, poster, abstract, grant submission, or presentation were categorized as follow-up analyses. Because the motivating scenarios were not known in advance, the eight categories of motivation (Table 1) were developed after the 100 consultations were completed. The eight categories were derived via consensus between Alison Callahan and Saurabh Gombar by reviewing the debriefs along with the PICOT form of the question, with Nigam H. Shah adjudicating differences. Naturally, these are not the entire universe of motivations possible, just the ones we observed in the requests we received.

### *Concordance of Consultation Results Across Data Sources*

We compared results obtained using different data sources for the same consultation request. To do so, we first identified consultations requesting treatment effect comparisons that could be re-executed using another data set. For example, if a consultation was originally completed using data from Stanford, we re-executed it using MarketScan and Optum claims data. Some two-way comparisons across data sets failed because of few patients in a given data set (our threshold was 100 patients), whereas for others, the matching procedure resulted in groups with no overlap in their propensity score distributions and thus were unsuitable for comparison.[52]

> *"The analysis found that at the cutoffs in use (procalcitonin >0.5), a positive test was not associated with a positive blood culture. This finding, combined with further analyses, informed an institutional protocol change: procalcitonin values are no longer used to inform ordering a blood culture when deciding whether to discontinue antibiotic therapy.*

**Table 1. Clinical Motivations and Subsequent Actions Taken by Requestors**

| Clinical Motivations | Total | Follow-Up Actions | | |
|---|---|---|---|---|
| | | Changed Patient Care | Guided Further Research | Follow-Up Analyses |
| Evaluating institutional patient management | 29 | 4 | 23 | 9 |
| Profiling outcomes of approved drugs | 29 | 0 | 15 | 4 |
| Profiling associations between laboratory test results and outcomes | 11 | 3 | 5 | 3 |
| Prognosis for understudied presentations | 4 | 1 | 3 | 1 |
| Treatment comparison for understudied populations | 4 | 0 | 3 | 0 |
| Profiling associations between diseases | 3 | 0 | 2 | 0 |
| Profiling rare disease presentations | 2 | 2 | 0 | 0 |
| Profiling outcomes of nonpharmacological treatments | 1 | 0 | 1 | 0 |
| **Total** | **83** | **10** | **52** | **17** |

Each of the 83 completed consultation requests was assigned a single motivation category and one or more follow-up action categories. The "Changed patient care" category applies to care delivery for a current or future patient, whereas the others apply to research-related programs or projects. Note: In 27 of 83 requests, the requestor took no action in response; thus, the sum of the actions taken can be less than the number of requests in a motivation category. Source: The authors.

Because a consultation to provide a treatment comparison could involve more than one outcome, we summarized concordance in terms of the number of outcomes, rather than the number of consultations. We evaluated the concordance of results for 59 outcomes from 33 consultations across Stanford and Optum and 53 outcomes from 22 consultations across Stanford and MarketScan.

Using the notion of agreement proposed for use in regulatory decisions,[53] a result was counted as concordant across two data sets only if both data sets provided an effect estimate in the same direction (e.g., both greater than 1 or both less than 1) with a $P$ value $\leq .05$, or if the effect estimates derived from both data sets did not indicate a significant effect on the outcome(s) of interest, regardless of direction.

## Findings from the First 100 Consultations

### *Consultations Requests Came from Multiple Specialties*

Of the first 100 requests sent over 18 months by 53 users from multiple specialties, 83 consultations were completed; 17 consultations could not be completed because of missing data

elements, available data sources having too few patients meeting the specified cohort criteria, inability to define a cohort, or the requirement for an unsupported study design. (These 17 are reflected in the failure mode designation in Figure 1.)

Of the 83 completed consultations, 48 were descriptive analyses; 35 were treatment comparison analyses, of which 18 had discrete or continuous outcomes and 17 had time-to-event outcomes; 78 of 83 (94%) consultations used Stanford EMR data, and four of 83 (5%) used national claims data to obtain adequate sample size. One consultation used both EMR and claims data.

Internal medicine was the most common requesting specialty, in terms of both requests received and number of requestors, followed by dermatology, oncology, and cardiology (Table 2). Among 53 users, 24 requested a consultation more than once, for a total of 76 consultations. Internal medicine also had the highest number of repeat users.

Median consultation turnaround time over the course of the 18-month pilot starting February 2, 2017, was 5 days, with 71 consultations (86%) completed in 10 days or less. Longer turnaround times occurred when additional data elements were needed, when there were delays in scheduling conversations with the requestor, or when matching required substantial time for large cohorts. As the service workflow matured, by the end of the study, 19 consultation reports were returned in 48 hours or less as a result of reusing cohort definitions, experience in PICOT formulation of the request, and analysis code optimization. This turnaround time is similar to that of send-out tests or e-consults at Stanford Health Care.

## *Consultation Requests Had Diverse Motivations*

Consultation requests were driven by a variety of motivations, including evaluating patient treatment strategies for a given disease or patient presentation, identifying comparatively effective treatments for patients with typically understudied characteristics, and quantifying associations between diseases. The categorization of consultation motivations is summarized in Table 1 and cross-tabulated with the subsequent actions taken by requestors: 10 consultations led to changes to patient care, 52 guided further research, and 17 led to follow-up analyses, including four that were presented at medical conferences or published in peer-reviewed journals.[54–57] Not all subsequent actions could be categorized into the three groups: 27 consultations lacked clear subsequent actions, suggesting that the consultation may have been sought primarily to contribute to the personal knowledge of the requestor or that the findings were not sufficiently compelling to warrant action on their basis.

In the Appendix, we highlight three consultations that demonstrate the diversity of situations motivating a consultation: a request to characterize a rare disease presentation (a pediatric patient with mononeuritis multiplex); a request to compare treatment outcomes (for a recently approved class of melanoma drugs, programmed death-1 [PD- 1] inhibitors); and a request to summarize the institutional use of procalcitonin tests (to inform antibiotic discontinuation). In each of these consultations, the service addressed a distinct need.

> *In the quest to make the best decisions for patients, clinicians use a hierarchy of evidence sources, and the report from our consultation service fits in as level III or IV in this hierarchy, because our consultations involved primarily case series and cohort studies.*

In the case of the patient with pediatric mononeuritis multiplex, the consultation required a custom descriptive analysis of a rare disease presentation that resulted in changes to patient care. We provided the requestor with summaries of the most frequent diagnoses preceding and following mononeuritis multiplex diagnosis in 118 similarly aged patients, which included bacterial and viral infections as well as psychosomatic disorders. A variety of treatments was

**Table 2. Summary of Completed Consultations by Specialty**

| | | Analysis Type | | |
|---|---|---|---|---|
| | | | Treatment Comparison | |
| Specialty | Consultations | Descriptive | Discrete or Continuous | Time-to-Event |
| Internal medicine | 21 | 10 | 11 | 0 |
| Dermatology | 9 | 8 | 1 | 0 |
| Oncology | 9 | 1 | 1 | 7 |
| Cardiology | 6 | 1 | 0 | 5 |
| General pediatrics | 5 | 5 | 0 | 0 |
| Emergency medicine | 4 | 2 | 2 | 0 |
| Endocrinology | 4 | 2 | 0 | 2 |
| Hematology | 4 | 1 | 1 | 2 |
| Allergy and immunology | 3 | 3 | 0 | 0 |
| Infectious disease | 3 | 3 | 0 | 0 |
| Pediatric neurology | 3 | 3 | 0 | 0 |
| Vascular surgery | 3 | 1 | 2 | 0 |
| Anesthesiology | 2 | 2 | 0 | 0 |
| Orthopedic surgery | 2 | 2 | 0 | 0 |
| Pathology | 2 | 2 | 0 | 0 |
| Nephrology | 1 | 1 | 0 | 0 |
| Ophthalmology | 1 | 0 | 0 | 1 |
| Urology | 1 | 1 | 0 | 0 |
| **Total** | **83** | **48** | **18** | **17** |

Internal medicine specialists made up about 25% of those requesting a consultation, followed by dermatology and oncology specialists, at about 11% each. There is some variation in the nature of the analysis types among and within specialties. For example, oncology and cardiology specialists more often sought time-to-event (survival) information. Source: The authors.

prescribed for those patients, including steroids, antibiotics, anti-inflammatory medications, painkillers, and hormone supplements. These findings, alongside further clinical workup, suggested managing the patient's symptoms as a postviral syndrome. The patient improved with a trial of steroids and was discharged.

In the case of the use of PD-1 inhibitors, the consultation required a treatment comparison analysis for an understudied population that guided further research. The consultation was motivated by a patient with melanoma who had a herpes simplex reactivation following treatment with nivolumab. We found 587 similar patients and found no difference in viral reactivation rates in patients treated with PD-1 inhibitors compared with those treated with other antineoplastic agents. Published evidence on the relationship between PD-1 therapies and herpetic reactivations was not available, perhaps because nivolumab was only recently approved (in 2014). In this case, the consultation findings filled an important clinical evidence gap.

In the case of procalcitonin testing, the request entailed a custom descriptive analysis to evaluate institutional patient management that changed patient care. The consultation was requested because while procalcitonin is a serum biomarker capable of discriminating between bacterial and nonbacterial causes of infection,[58–61] the exact cutoff value at which to discontinue antibiotics is not universally agreed upon. Procalcitonin's utility for deciding whether to order a blood culture remains unclear.[62,63] By analyzing approximately 16,000 procalcitonin test results and 29,000 blood culture results, we calculated how often a positive blood culture was obtained within 48 hours of one cutoff value for a procalcitonin result, how frequently antibiotic therapy was discontinued at different cutoffs of procalcitonin values, and how often antibiotics were restarted within 72 hours of discontinuation. The analysis found that at the cutoffs in use (procalcitonin >0.5), a positive test was not associated with a positive blood culture. This finding, combined with further analyses, informed an institutional protocol change: procalcitonin values are no longer used to inform ordering a blood culture when deciding whether to discontinue antibiotic therapy.

### *Consultation Results Were Concordant Across Data Sets*

When comparing results obtained using different data sources for the same consultation request, 68% to 74% of results were concordant across data sets. In the Stanford and Optum comparison, results for 68% of the evaluated outcomes (40 of 59) were concordant. For 28 outcomes, both data sets reported a significant treatment effect with the same direction of the effect. For 12 outcomes, results from both data sets did not have a significant effect. In the Stanford and MarketScan comparison, 74% of the evaluated outcomes (39 of 53) were concordant. For 30 outcomes, results from both data sets had a significant treatment effect with the same direction of the effect. For nine outcomes, results from both data sets did not have a significant effect.

## A Vision Realized: Strengths, Caveats, and Next Steps

Using data generated during routine care to guide the personalized care of patients — both for individuals during a given health system encounter and for subsequent patients with similar

circumstances — is a core tenet of a learning health system[64–67] and, as a distillation of clinical expertise, of evidence-based medicine.[1,68] Our work is an implementation of this vision, demonstrating that an on-demand consultation service to summarize the experiences of inpatients and previously seen patients is feasible from both an engineering and an operational standpoint. The variety of consultation requests we received (in terms of clinical motivations, analyses needed, and subsequent actions) also empirically illustrates the potential to inform a broad range of clinical decisions (Figure 2).

As large patient data repositories become commonplace,[69,70] the ability to learn from the experience of similar patients is one of the nobler opportunities such repositories enable.[71] In the quest to make the best decisions for patients, clinicians use a hierarchy of evidence sources,[72] and the report from our consultation service fits in as level III or IV in this hierarchy, because our consultations involved primarily case series and cohort studies. It may also be possible to combine the report from a service such as ours with clinical librarian services[73] to integrate literature-derived evidence with learning from past patient data.

> *We found 68%–74% concordance of consultation results across multiple data sets, a rate of agreement comparable to both the rate at which results from RCTs agree with each other (67%–87%) and the rate at which results derived from observational claims data agree with RCTs (60%–80%).*

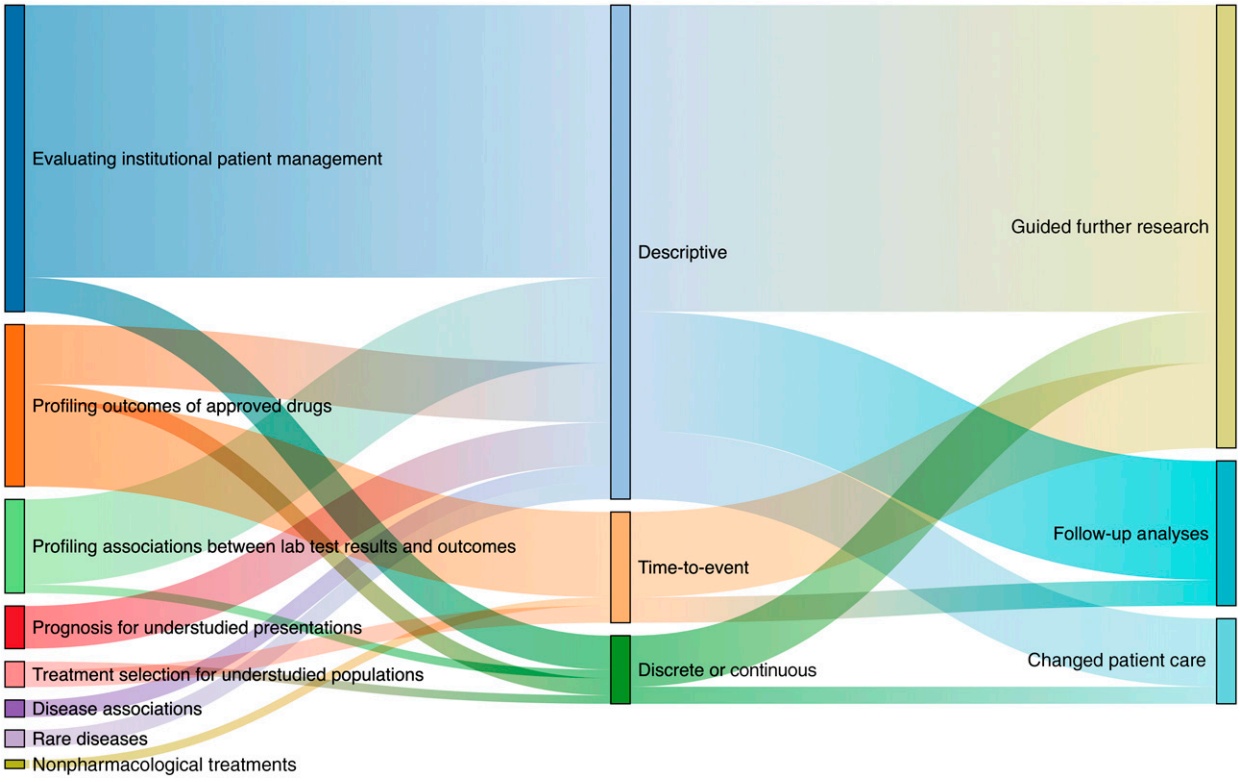
Our work has several unique strengths. First, the service's underlying search engine, ACE — which is essential for rapidly constructing cohorts and defining electronic phenotypes corresponding to exposures and outcomes of interest — as well as the analysis code are available via a complimentary license for noncommercial use, allowing implementation of such a service at other academic sites without a fee;[18] software licensing is managed by the Stanford Office of Technology Licensing. Second, we found 68%–74% concordance of consultation results across multiple data sets, a rate of agreement comparable to both the rate at which results from RCTs agree with each other (67%–87%)[74] and the rate at which results derived from observational claims data agree with RCTs (60%–80%).[75] Finally, the number of repeat requests demonstrates the need for, as well as the viability of, such a consultation service; the rate at which the results changed patient care (12%) is within the range at which an expert second opinion impacts care (10%–62%), albeit at the lower end.[76]

Our study has several limitations. First, we only responded to 100 consultations from 53 users who were self-selecting; consultation requestors may thus have been predisposed to finding value in the service, and self-reported utility for advancing research or patient care may have been affected by subjective expectations of the service. Second, the cost to deploy such a service will vary at institutions where the necessary data access and analysis infrastructure do not yet exist. Our implementation was at an academic medical center with access to personnel experienced in analyzing messy EMR data as well as ready access to EMR and claims data,

FIGURE 2.

## Use Cases that Can Be Served by an On-Demand Consultation Service in a Learning Health System

This figure uses a Sankey plot to illustrate the flow (horizontal colored lines) of completed consultations in terms of the clinical motivation (left) to the analysis type (center) to the subsequent action (right). The thickness of each flow is proportional to the number of consultations. For example, consultations motivated by evaluating institutional patient management required mostly descriptive analyses and resulted in all three categories of subsequent action, including changing patient care.




Source: The authors

NEJM Catalyst (catalyst.nejm.org) © Massachusetts Medical Society

resulting in an estimated cost per consult at $550 U.S. dollars;[14,16] costs may be higher elsewhere. Moreover, while the current turnaround time is analogous to a send-out laboratory test, providing an ongoing service would require additional engineering effort, incurring additional costs.[77] Third, the choice and evaluation of patient matching and causal inference methods remains an active area of research.[78–80] Future work may find methods beyond hd-PSM that offer improved concordance across data sources. In hindsight, using just the hd-PSM would have resulted in a simpler report. In addition, estimating the error rate in the absence of ground truth for the questions received remains difficult. The 17%–20% false-positive rate is applicable to the 35 consultations requesting a comparison between two treatments because the "ground

truth" in the OMOP and EU-ADR reference sets is designed for such comparisons. Finally, the net benefit of providing on-demand evidence needs to be studied prospectively at multiple sites by measuring the long-term impact on patient outcomes, cost of care, and health system operations, which we were unable to do in this study. We hope that our experience and the tooling we share will enable such studies.

## Takeaways

On-demand evidence generation to inform clinical decision-making is an achievable goal, given the confluence of scalable technology for data analysis, a growing data science workforce, the training of increasingly data-savvy clinicians, and the availability of large amounts of patient data from EMRs and claims.[8,14] The consultation service we created provides proof-of-feasibility for realizing this goal. Such a service is capable of informing patient care at the bedside for specific patients, informing the medical literature, and supporting institutional guideline creation. As large patient data repositories are created,[75,77] the potential to benefit from such a service is immense.[78] Given the feasibility and the documented need, studies establishing the utility of having such a consultation service are logical next steps.

**Alison Callahan, PhD, MIS**
Research Scientist, Center for Biomedical Informatics, Department of Medicine, School of Medicine, Stanford University, Stanford, California, USA

**Saurabh Gombar, MD, PhD**
Instructor, Department of Pathology, School of Medicine, Stanford University, Stanford, California, USA

Chief Medical Officer, Atropos Health, Palo Alto, California, USA

**Eli M. Cahan**
Research Student, Knight-Hennessy Scholar, Department of Medicine, School of Medicine, Stanford University, Stanford, California, USA

Medical Student, New York University School of Medicine, New York, New York, USA

**Kenneth Jung, PhD**
Research Scientist, Department of Medicine, School of Medicine, Stanford University, Stanford, California, USA

**Ethan Steinberg**
PhD Student, Department of Computer Science, Stanford University, Stanford, California, USA

**Vladimir Polony**
Senior Research Engineer, Department of Medicine, School of Medicine, Stanford University, Stanford, California, USA

Director of Engineering, Atropos Health, Palo Alto, California, USA

**Keith Morse, MD, MBA**
Medical Director of Clinical Informatics, Stanford Children's Health, Stanford, California, USA

Clinical Assistant Professor, Department of Pediatrics (Hospital Medicine), School of Medicine, Stanford University, Stanford, California, USA

**Robert Tibshirani, PhD, MS**
Professor of Biomedical Data Science and Statistics, Department of Statistics, School of Humanities & Sciences, Stanford University, Stanford, California, USA

**Trevor Hastie, PhD, MS**
Professor of Statistics and of Biomedical Data Sciences, Department of Statistics, School of Humanities & Sciences, Stanford University, Stanford, California, USA

**Robert Harrington, MD**
Professor of Medicine and Chairman, Department of Medicine, School of Medicine, Stanford University, Stanford, California, USA

**Nigam H. Shah, PhD, MBBS**
Professor of Medicine and Biomedical Data Science, School of Medicine, Stanford University, Stanford, California, USA

Associate Chief Information Officer for Data Science, Stanford Healthcare, Palo Alto, California, USA

## Appendix

Example Consultation Reports and Structured Debrief Script

## Acknowledgments

We thank Dr. Christopher Sharp for guidance on setting up the effort as a referral service, Michael Halaas for guidance on the computational environment setup, Tip Kim for guidance on the licensing setup, and David Entwistle for supporting the vision of scaling such an effort beyond Stanford Health Care.

## Disclosures

The study was supported by an anonymous philanthropic gift, Grant 5R01LM011369 from the National Library of Medicine, and institutional support from the School of Medicine, Department of Medicine, and Stanford Health Care. Saurabh Gombar serves as the Chief Medical Officer and Vladimir Polony serves as Director of Engineering for Atropos Health, which offers a clinician-backed digital consultation via proprietary technology developed at

Stanford. The company was launched in October 2020 with support from Stanford Health Care's Innovation Program. Alison Callahan and Nigam H. Shah are technical advisors to Atropos Health. Eli M. Cahan, Kenneth Jung, Ethan Steinberg, Keith Morse, Robert Tibshirani, Trevor Hastie, and Robert Harrington have nothing to disclose.

## References


1. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ 1996;312:71-2 https://www.bmj.com/content/312/7023/71 https://doi.org/10.1136/bmj.312.7023.71.

2. Institute of Medicine (US) Committee on Quality of Health Care in America. To Err is Human: Building a Safer Health System. Washington, DC: National Academies Press, 2000. https://pubmed.ncbi.nlm.nih.gov/25077248/. https://doi.org/10.17226/9728.

3. Makary MA, Daniel M. Medical error-the third leading cause of death in the US. BMJ 2016;353:i2139 https://www.bmj.com/content/353/bmj.i2139 https://doi.org/10.1136/bmj.i2139.

4. Rogers JR, Liu C, Hripcsak G, Cheung YK, Weng C. Comparison of clinical characteristics between clinical trial participants and nonparticipants using electronic health record data. JAMA Netw Open 2021;4:e214732 https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2778202 https://doi.org/10.1001/jamanetworkopen.2021.4732.

5. Black N. Why we need observational studies to evaluate the effectiveness of health care. BMJ 1996; 312:1215-8 https://www.bmj.com/content/312/7040/1215 https://doi.org/10.1136/bmj.312.7040.1215.

6. Stewart WF, Shah NR, Selna MJ, Paulus RA, Walker JM. Bridging the inferential gap: the electronic health record and clinical evidence. Health Aff (Millwood) 2007;26(Suppl 1):w181-91 https://www.healthaffairs.org/doi/10.1377/hlthaff.26.2.w181 https://doi.org/10.1377/hlthaff.26.2.w181.

7. Califf RM, Robb MA, Bindman AB, et al. Transforming evidence generation to support health and health care decisions. N Engl J Med 2016;375:2395-400 https://www.nejm.org/doi/10.1056/NEJMsb1610128 https://doi.org/10.1056/NEJMsb1610128.

8. Longhurst CA, Harrington RA, Shah NHA. A 'green button' for using aggregate patient data at the point of care. Health Aff (Millwood) 2014;33:1229-35 https://www.healthaffairs.org/doi/10.1377/hlthaff.2014.0099 https://doi.org/10.1377/hlthaff.2014.0099.

9. Institute of Medicine (US) Roundtable on Evidence-Based Medicine. The Learning Healthcare System: Workshop Summary. Washington, DC: National Academies Press, 2007. https://www.nap.edu/catalog/11903/the-learning-healthcare-system-workshop-summary. https://doi.org/10.17226/11903.

10. Feinstein AR, Rubinstein JF, Ramshaw WA. Estimating prognosis with the aid of a conversational-mode computer program. Ann Intern Med 1972;76:911-21 https://www.acpjournals.org/doi/10.7326/0003-4819-76-6-911 https://doi.org/10.7326/0003-4819-76-6-911.

11. Rosati RA, McNeer JF, Starmer CF, Mittler BS, Morris JJ Jr, Wallace AG. A new information system for medical practice. Arch Intern Med 1975;135:1017-24 https://jamanetwork.com/journals/jamainternalmedicine/article-abstract/584728 https://doi.org/10.1001/archinte.1975.00330080019003.

12. Frankovich J, Longhurst CA, Sutherland SM. Evidence-based medicine in the EMR era. N Engl J Med 2011;365:1758-9 https://www.nejm.org/doi/full/10.1056/NEJMp1108726 https://doi.org/10.1056/NEJMp1108726.

13. Institute of Medicine. Digital Infrastructure for the Learning Health System: The Foundation for Continuous Improvement in Health and Health Care: Workshop Series Summary. Washington, DC: National Academies Press, 2011. https://www.ncbi.nlm.nih.gov/books/NBK83569/.

14. Gombar S, Callahan A, Califf R, Harrington R, Shah NH. It is time to learn from patients like mine. NPJ Digit Med 2019;2:16 https://www.nature.com/articles/s41746-019-0091-3.

15. Callahan A, Shah NH, Chen JH. Research and reporting considerations for observational studies using electronic health record data. Ann Intern Med 2020;172(Suppl):S79-84 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7413106/ https://doi.org/10.7326/M19-0873.

16. Schuler A, Callahan A, Jung K, Shah NH. Performing an informatics consult: methods and challenges. J Am Coll Radiol 2018;15(3 Pt B):563-8 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5901653/ https://doi.org/10.1016/j.jacr.2017.12.023.

17. Datta S, Posada J, Olson G, et al. A new paradigm for accelerating clinical data science at Stanford Medicine. arXiv March 17, 2020 [Online ahead of print] https://arxiv.org/abs/2003.10534.

18. Callahan A, Polony V, Posada JD, Banda JM, Gombar S, Shah NH. ACE: the Advanced Cohort Engine for searching longitudinal patient records. J Am Med Inform Assoc 2021;28:1468-79 https://academic.oup.com/jamia/article/28/7/1468/6169466 https://doi.org/10.1093/jamia/ocab027.

19. Wang G, Jung K, Winnenburg R, Shah NH. A method for systematic discovery of adverse drug events from clinical notes. J Am Med Inform Assoc 2015;22:1196-204 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4921953/ https://doi.org/10.1093/jamia/ocv102.

20. LePendu P, Iyer SV, Bauer-Mehren A, et al. Pharmacovigilance using clinical notes. Clin Pharmacol Ther 2013;93:547-55 https://ascpt.onlinelibrary.wiley.com/doi/10.1038/clpt.2013.47 https://doi.org/10.1038/clpt.2013.47.

21. Callahan A, Fries JA, Ré C, et al. Medical device surveillance with electronic health records. NPJ Digit Med 2019;2:94 https://www.nature.com/articles/s41746-019-0168-z. https://doi.org/10.1038/s41746-019-0168-z.

22. Nead KT, Gaskin G, Chester C, et al. Androgen deprivation therapy and future Alzheimer's disease risk. J Clin Oncol 2016;34:566-71 https://ascopubs.org/doi/10.1200/JCO.2015.63.6266 https://doi.org/10.1200/JCO.2015.63.6266.

23. Nead KT, Gaskin G, Chester C, Swisher-McClure S, Leeper NJ, Shah NH. Association between androgen deprivation therapy and risk of dementia. JAMA Oncol 2017;3:49-55 https://jamanetwork.com/journals/jamaoncology/fullarticle/2569059 https://doi.org/10.1001/jamaoncol.2016.3662.

24. Tamang S, Patel MI, Blayney DW, et al. Detecting unplanned care from clinician notes in electronic health records. J Oncol Pract 2015;11:e313-9 https://ascopubs.org/doi/10.1200/JOP.2014.002741 https://doi.org/10.1200/JOP.2014.002741.

25. Hernandez-Boussard T, Tamang S, Blayney D, Brooks J, Shah N. New paradigms for patient-centered outcomes research in electronic medical records: an example of detecting urinary incontinence following prostatectomy. EGEMS (Wash DC) 2016;4:1231 https://egems.academyhealth.org/articles/abstract/10.13063/2327-9214.1231/. https://doi.org/10.13063/2327-9214.1231.

26. Caswell-Jin JL, Callahan A, Purington N, et al. Treatment and monitoring variability in US metastatic breast cancer care. JCO Clin Cancer Inform 2021;5:600-14 https://ascopubs.org/doi/abs/10.1200/CCI.21.00031 https://doi.org/10.1200/CCI.21.00031.

27. Vashisht R, Jung K, Schuler A, et al. Association of hemoglobin A1c levels with use of sulfonylureas, dipeptidyl peptidase 4 inhibitors, and thiazolidinediones in patients with type 2 diabetes treated with metformin: analysis from the Observational Health Data Sciences and Informatics Initiative. JAMA Netw Open 2018;1:e181755 https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2698083 https://doi.org/10.1001/jamanetworkopen.2018.1755.

28. Banda JM, Seneviratne M, Hernandez-Boussard T, Shah NH. Advances in electronic phenotyping: from rule-based definitions to machine learning models. Annu Rev Biomed Data Sci 2018;1:53-68 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6583807/ https://doi.org/10.1146/annurev-biodatasci-080917-013315.

29. Seneviratne MG, Banda JM, Brooks JD, Shah NH, Hernandez-Boussard TM. Identifying cases of metastatic prostate cancer using machine learning on electronic health records. AMIA Annu Symp Proc 2018;2018:1498-504 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6371284/.

30. Agarwal V, Podchiyska T, Banda JM, et al. Learning statistical models of phenotypes using noisy labeled training data. J Am Med Inform Assoc 2016;23:1166-73 https://academic.oup.com/jamia/article/23/6/1166/2399304 https://doi.org/10.1093/jamia/ocw028.

31. Banda JM, Halpern Y, Sontag D, Shah NH. Electronic phenotyping with APHRODITE and the Observational Health Sciences and Informatics (OHDSI) data network. AMIA Jt Summits Transl Sci Proc 2017;2017:48-57 https://pubmed.ncbi.nlm.nih.gov/28815104/.

32. Kim Y, Tian Y, Yang J, et al. Comparative safety and effectiveness of alendronate versus raloxifene in women with osteoporosis. Sci Rep 2020;10:11115 https://www.nature.com/articles/s41598-020-68037-8. https://doi.org/10.1038/s41598-020-68037-8.

33. Chen R, Ryan P, Natarajan K, et al. Treatment patterns for chronic comorbid conditions in patients with cancer using a large-scale observational data network. JCO Clin Cancer Inform 2020;4:171-83 https://ascopubs.org/doi/10.1200/CCI.19.00107 https://doi.org/10.1200/CCI.19.00107.

34. Prats-Uribe A, Sena AG, Lai LYH, et al. Use of repurposed and adjuvant drugs in hospital patients with covid-19: multinational network cohort study. BMJ 2021;373:n1038 https://www.bmj.com/content/373/bmj.n1038.

35. Tan EH, Sena AG, Prats-Uribe A, et al. COVID-19 in patients with autoimmune diseases: characteristics and outcomes in a multinational network of cohorts across three countries. Rheumatology (Oxford) March 16, 2021 [Online ahead of print] https://academic.oup.com/rheumatology/advance-article/doi/10.1093/rheumatology/keab250/6174122. https://doi.org/10.1093/rheumatology/keab250.

36. Burn E, You SC, Sena AG, et al. Deep phenotyping of 34,128 adult patients hospitalised with COVID-19 in an international network study. Nat Commun 2020;11:5009 https://www.nature.com/articles/s41467-020-18849-z https://doi.org/10.1038/s41467-020-18849-z.

37. Duke JD, Ryan PB, Suchard MA, et al. Risk of angioedema associated with levetiracetam compared with phenytoin: findings of the observational health data sciences and informatics research network. Epilepsia 2017;58:e101-6 https://onlinelibrary.wiley.com/doi/10.1111/epi.13828 https://doi.org/10.1111/epi.13828.

38. Hripcsak G, Ryan PB, Duke JD, et al. Characterizing treatment pathways at scale using the OHDSI network. Proc Natl Acad Sci USA 2016;113:7329-36 https://www.pnas.org/content/113/27/7329 https://doi.org/10.1073/pnas.1510502113.

39. Lowe HJ, Ferris TA, Hernandez PM, Weber SC. STRIDE—an integrated standards-based translational research informatics platform. AMIA Annu Symp Proc 2009;2009:391-5 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2815452/.

40. Lependu P, Iyer SV, Fairon C, Shah NH. Annotation analysis for testing drug safety signals using unstructured clinical notes. J Biomed Semantics 2012;3(Suppl 1):S5 https://jbiomedsem.biomedcentral.com/articles/10.1186/2041-1480-3-S1-S5 https://doi.org/10.1186/2041-1480-3-S1-S5.

41. Lependu P, Liu Y, Iyer S, Udell MR, Shah NH. Analyzing patterns of drug use in clinical notes for patient safety. AMIA Jt Summits Transl Sci Proc 2012;2012:63-70 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3392046/.

42. Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. Am J Epidemiol 2016;183:758-64 https://academic.oup.com/aje/article/183/8/758/1739860 https://doi.org/10.1093/aje/kwv254.

43. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivariate Behav Res 2011;46:399-424 https://www.tandfonline.com/doi/full/10.1080/00273171.2011.568786 https://doi.org/10.1080/00273171.2011.568786.

44. Guertin JR, Rahme E, Dormuth CR, LeLorier J. Head to head comparison of the propensity score and the high-dimensional propensity score matching methods. BMC Med Res Methodol 2016;16:22 https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-016-0119-1 https://doi.org/10.1186/s12874-016-0119-1.

45. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw 2010;33:1-22 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2929880/ https://doi.org/10.18637/jss.v033.i01.

46. Stuart EA. Matching methods for causal inference: a review and a look forward. Stat Sci 2010;25:1-21 https://projecteuclid.org/journals/statistical-science/volume-25/issue-1/Matching-Methods-for-Causal-Inference-A-Review-and-a/10.1214/09-STS313.full https://doi.org/10.1214/09-STS313.

47. Ryan PB, Schuemie MJ, Welebob E, Duke J, Valentine S, Hartzema AG. Defining a reference set to support methodological research in drug safety. Drug Saf 2013;36(Suppl 1):S33-47 https://link.springer.com/article/10.1007/s40264-013-0097-8 https://doi.org/10.1007/s40264-013-0097-8.

48. Coloma PM, Avillach P, Salvo F, et al. A reference standard for evaluation of methods for drug safety signal detection using electronic healthcare record databases. Drug Saf 2013;36:13-23 https://link.springer.com/article/10.1007/s40264-012-0002-x https://doi.org/10.1007/s40264-012-0002-x.

49. Eriksson R, Werge T, Jensen LJ, Brunak S. Dose-specific adverse drug reaction identification in electronic patient records: temporal data mining in an inpatient psychiatric population [published correction appears in Drug Saf 2014;37:379]. Drug Saf 2014;37:237-47 https://link.springer.com/article/10.1007%2Fs40264-014-0145-z https://doi.org/10.1007/s40264-014-0145-z.

50. Xiao C, Li Y, Baytas IM, Zhou J, Wang F. An MCEM framework for drug safety signal detection and combination from heterogeneous real world evidence. Sci Rep 2018;8:1806 https://www.nature.com/articles/s41598-018-19979-7.

51. Ryan PB, Buse JB, Schuemie MJ, et al. Comparative effectiveness of canagliflozin, SGLT2 inhibitors and non-SGLT2 inhibitors on the risk of hospitalization for heart failure and amputation in patients with type 2 diabetes mellitus: a real-world meta-analysis of 4 observational databases (OBSERVE-4D). Diabetes Obes Metab 2018;20:2585-97 https://dom-pubs.onlinelibrary.wiley.com/doi/10.1111/dom.13424 https://doi.org/10.1111/dom.13424.

52. Oberst M, Johansson FD, Wei D, et al. Characterization of overlap in observational studies. Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics, PMLR 2020; 108:788-798.

53. Franklin JM, Pawar A, Martin D, et al. Nonrandomized real-world evidence to support regulatory decision making: process for a randomized trial replication project. Clin Pharmacol Ther 2020;107:817-26 https://ascpt.onlinelibrary.wiley.com/doi/10.1002/cpt.1633 https://doi.org/10.1002/cpt.1633.

54. Rhee J, Gombar S, Beutner K, Callahan A, Jung K, Wheeler M. Cardiovascular safety of mexiletine as a therapy for myotonia in patients with myotonic dystrophy. J Am Coll Cardiol 2019;73(Suppl 1):

474 https://www.jacc.org/doi/full/10.1016/S0735-1097%2819%2931082-4 https://doi.org/10.1016/S0735-1097(19)31082-4.

55. Karimi Y, Gombar S, Dean L, et al. Real-world efficacy of bone modifying agents (BMAs) in patients with breast cancer (BC) treated in an academic health system: use of the "green button. J Clin Oncol 2019;37(Suppl):e18054 https://ascopubs.org/doi/abs/10.1200/JCO.2019.37.15_suppl.e18054 https://doi.org/10.1200/JCO.2019.37.15_suppl.e18054.

56. Meng L, Gombar S, Callahan A, et al. 210. Step-down from IV to oral therapy in patients with bacteremia due to Enterobacteriaceae: fluoroquinolones (FQ) vs. ß-lactams (BL) or trimethoprim-sulfamethoxazole (TMP-SMX). Open Forum Infect Dis 2019;6(Suppl 2):S124 https://academic.oup.com/ofid/article/6/Supplement_2/S124/5604412 https://doi.org/10.1093/ofid/ofz360.285.

57. Ibrahim B, de Freitas Mendonca MI, Gombar S, Callahan A, Jung K, Capasso R. Association of systemic diseases with surgical treatment for obstructive sleep apnea compared with continuous positive airway pressure. JAMA Otolaryngol Head Neck Surg 2021;147(4):329–335 doi: 10.1001/jamaoto.2020.5179.

58. Christ-Crain M, Jaccard-Stolz D, Bingisser R, et al. Effect of procalcitonin-guided treatment on antibiotic use and outcome in lower respiratory tract infections: cluster-randomised, single-blinded intervention trial. Lancet 2004;363:600-7 https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(04)15591-8/fulltext https://doi.org/10.1016/S0140-6736(04)15591-8.

59. Schuetz P, Christ-Crain M, Thomann R, et al., ProHOSP Study Group. Effect of procalcitonin-based guidelines vs standard guidelines on antibiotic use in lower respiratory tract infections: the ProHOSP randomized controlled trial. JAMA 2009;302:1059-66 https://jamanetwork.com/journals/jama/fullarticle/184544 https://doi.org/10.1001/jama.2009.1297.

60. Tonkin-Crine SK, Tan PS, van Hecke O, et al. Clinician-targeted interventions to influence antibiotic prescribing behaviour for acute respiratory infections in primary care: an overview of systematic reviews. Cochrane Database Syst Rev 2017;9:CD012252 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6483738/ https://doi.org/10.1002/14651858.CD012252.pub2.

61. Schuetz P, Wirz Y, Sager R, et al. Procalcitonin to initiate or discontinue antibiotics in acute respiratory tract infections. Cochrane Database Syst Rev 2017;10:CD007498 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6485408/ https://doi.org/10.1002/14651858.CD007498.pub3.

62. Wu Q, Yang H, Kang Y. Comparison of diagnostic accuracy among procalcitonin, C-reactive protein, and interleukin 6 for blood culture positivity in general ICU patients. Crit Care 2018;22:339 https://ccforum.biomedcentral.com/articles/10.1186/s13054-018-2269-5.

63. Bassetti M, Russo A, Righi E, et al. Comparison between procalcitonin and C-reactive protein to predict blood culture results in ICU patients. Crit Care 2018;22:252 https://ccforum.biomedcentral.com/articles/10.1186/s13054-018-2183-x.

64. Grumbach K, Lucey CR, Johnston SC. Transforming from centers of learning to learning health systems: the challenge for academic health centers. JAMA 2014;311:1109-10 https://jamanetwork.com/journals/jama/article-abstract/1841977 https://doi.org/10.1001/jama.2014.705.

65. Krumholz HM, Terry SF, Waldstreicher J. Data acquisition, curation, and use for a continuously learning health system. JAMA 2016;316:1669-70 https://jamanetwork.com/journals/jama/article-abstract/2556006 https://doi.org/10.1001/jama.2016.12537.

66. Smoyer WE, Embi PJ, Moffatt-Bruce S. Creating local learning health systems: think globally, act locally. JAMA 2016;316:2481-2 http://hbiostat.org/papers/medical/learningHealthSystem/smo16cre.pdf.

67. Bindman AB, Pronovost PJ, Asch DA. Funding innovation in a learning health care system. JAMA 2018;319:119-20 https://jamanetwork.com/journals/jama/article-abstract/2667095 https://doi.org/10.1001/jama.2017.18205.

68. Evidence-Based Medicine Working Group. Evidence-based medicine. A new approach to teaching the practice of medicine. JAMA 1992;268:2420-5 https://jamanetwork.com/journals/jama/article-abstract/400956 https://doi.org/10.1001/jama.1992.03490170092032.

69. Mandl KD, Kohane IS. Tectonic shifts in the health information economy. N Engl J Med 2008;358:1732-7 https://www.nejm.org/doi/full/10.1056/NEJMsb0800220 https://doi.org/10.1056/NEJMsb0800220.

70. Mandl KD, Perakslis ED. HIPAA and the leak of "deidentified" EHR data. N Engl J Med 2021;384:2171-3 https://www.nejm.org/doi/full/10.1056/NEJMp2102616.

71. Palmer K. 'Some very noble, some less than noble': the growing health data marketplace sparks concern over patient privacy. STAT. June 9, 2021. Accessed June 15, 2021. https://www.statnews.com/2021/06/09/datavant-ciox-health-data-hipaa/.

72. Burns PB, Rohrich RJ, Chung KC. The levels of evidence and their role in evidence-based medicine. Plast Reconstr Surg 2011;128:305-10 https://journals.lww.com/plasreconsurg/Fulltext/2011/07000/The_Levels_of_Evidence_and_Their_Role_in.46.aspx https://doi.org/10.1097/PRS.0b013e318219c171.

73. Mulvaney SA, Bickman L, Giuse NB, Lambert EW, Sathe NA, Jerome RN. A randomized effectiveness trial of a clinical informatics consult service: impact on evidence-based decision-making and knowledge implementation. J Am Med Inform Assoc 2008;15:203-11 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2274783/ https://doi.org/10.1197/jamia.M2461.

74. Ryan P, Hripcsak G. A journey toward real-world evidence for regulatory decision-making: proving reliable real-world evidence: Replicating RCTs using LEGEND. Observation Health Data Sciences and Informatics. 2019. Accessed August 12, 2021. https://www.ohdsi.org/wp-content/uploads/2019/09/4-Plenary-3-Replicating-LEGEND.pdf.

75. Franklin JM, Patorno E, Desai RJ, et al. Emulating randomized clinical trials with nonrandomized real-world evidence studies: first results from the RCT DUPLICATE Initiative. Circulation 2021;143:1002-13 https://www.ahajournals.org/doi/10.1161/CIRCULATIONAHA.120.051718.

76. Payne VL, Singh H, Meyer AND, Levy L, Harrison D, Graber ML. Patient-initiated second opinions: systematic review of characteristics and impact on diagnosis, treatment, and satisfaction. Mayo Clin Proc 2014;89:687-96 https://www.mayoclinicproceedings.org/article/S0025-6196(14)00245-6/pdf https://doi.org/10.1016/j.mayocp.2014.02.015.

77. Norgeot B, Glicksberg BS, Butte AJ. A call for deep-learning healthcare. Nat Med 2019;25:14-5 https://www.nature.com/articles/s41591-018-0320-3 https://doi.org/10.1038/s41591-018-0320-3.

78. King G, Lucas C, Nielsen RA. The balance-sample size frontier in matching methods for causal inference. Am J Pol Sci 2017;61:473-89 https://onlinelibrary.wiley.com/doi/10.1111/ajps.12272 https://doi.org/10.1111/ajps.12272.

79. Powers S, Qian J, Jung K, et al. Some methods for heterogeneous treatment effect estimation in high dimensions. Stat Med 2018;37:1767-87 https://onlinelibrary.wiley.com/doi/10.1002/sim.7623 https://doi.org/10.1002/sim.7623.

80. Liu L, Mukherjee R, Robins JM. On nearly assumption-free tests of nominal confidence interval coverage for causal parameters estimated by machine learning. arXiv July 12, 2020 [Online ahead of print] https://arxiv.org/pdf/1904.04276.pdf.