

# Rapid Development of Specialty Population Registries and Quality Measures from Electronic Health Record Data\*

## An Agile Framework

Vaishnavi Kannan; Jason S. Fish; Jacqueline M. Mutz; Angela R. Carrington; Ki Lai; Lisa S. Davis; Josh E. Youngblood; Mark R. Rauschuber; Kathryn A. Flores; Evan J. Sara; Deepa G. Bhat; DuWayne L. Willett

University of Texas Southwestern Medical Center, Dallas, TX, USA

### Keywords

Registries, electronic health records, data collection, outcome and process assessment (health care), quality indicators, information storage and retrieval, data warehouse, agile development, population health

### Summary

**Background:** Creation of a new electronic health record (EHR)-based registry often can be a “one-off” complex endeavor: first developing new EHR data collection and clinical decision support tools, followed by developing registry-specific data extractions from the EHR for analysis. Each development phase typically has its own long development and testing time, leading to a prolonged overall cycle time for delivering one functioning registry with companion reporting into production. The next registry request then starts from scratch. Such an approach will not scale to meet the emerging demand for specialty registries to support population health and value-based care.

**Objective:** To determine if the creation of EHR-based specialty registries could be markedly accelerated by employing (a) a finite core set of EHR data collection principles and methods, (b) concurrent engineering of data extraction and data warehouse design using a common dimensional data model for all registries, and (c) agile development methods commonly employed in new product development.

**Methods:** We adopted as guiding principles to (a) capture data as a byproduct of care of the patient, (b) reinforce optimal EHR use by clinicians, (c) employ a finite but robust set of EHR data capture tool types, and (d) leverage our existing technology toolkit. Registries were defined by a shared condition (recorded on the Problem List) or a shared exposure to a procedure (recorded on the Surgical History) or to a medication (recorded on the Medication List). Any EHR fields needed – either to determine registry membership or to calculate a registry-associated clinical quality measure (CQM) – were included in the enterprise data warehouse (EDW) shared dimensional data

model. Extract-transform-load (ETL) code was written to pull data at defined “grains” from the EHR into the EDW model. All calculated CQM values were stored in a single Fact table in the EDW crossing all registries. Registry-specific dashboards were created in the EHR to display both (a) real-time patient lists of registry patients and (b) EDW-generated CQM data. Agile project management methods were employed, including co-development, lightweight requirements documentation with User Stories and acceptance criteria, and time-boxed iterative development of EHR features in 2-week “sprints” for rapid-cycle feedback and refinement.

**Results:** Using this approach, in calendar year 2015 we developed a total of 43 specialty chronic disease registries, with 111 new EHR data collection and clinical decision support tools, 163 new clinical quality measures, and 30 clinic-specific dashboards reporting on both real-time patient care gaps and summarized and vetted CQM measure performance trends.

**Conclusions:** This study suggests concurrent design of EHR data collection tools and reporting can quickly yield useful EHR structured data for chronic disease registries, and bodes well for efforts to migrate away from manual abstraction. This work also supports the view that in new EHR-based registry development, as in new product development, adopting agile principles and practices can help deliver valued, high-quality features early and often.

### Correspondence to:

DuWayne L. Willett, MD, MS  
Professor, Department of Internal Medicine  
Chief Medical Informatics Officer  
University of Texas Southwestern Health System  
5323 Harry Hines Blvd.  
Dallas, TX 75390-9047  
USA  
E-mail: DuWayne.Willett@UTSouthwestern.edu

Methods Inf Med 2017; 56(Open): e74–e83

<https://doi.org/10.3414/ME16-02-0031>

received: September 15, 2016

accepted: April 19, 2017

published: June 14, 2017

\* Supplementary material published on our website <https://doi.org/10.3414/ME16-02-0031>

## 1. Introduction

### 1.1 Overview

Can clinical data entered in now-ubiquitous electronic health records (EHRs) drive creation of new specialty patient registries and quality measures? With the move towards value-based care [1], specialty registries are increasingly needed for measuring and improving the health of detailed subpopulations of patients [2, 3, 4, 5, 6]. Manual abstraction of medical record data quickly becomes cost-prohibitive, and doesn't scale. Prior EHR-based registry development projects at our institution were each "one-off" projects, with a long development time for the EHR data collection tools, followed by a long development time for an idiosyncratic registry data extraction and reporting capability. 12 months end-to-end elapsed project time was typical. Yet in 2015 we were given the challenge of developing at least one specialty registry with associated clinical quality measures (CQMs) for each specialty in our academic medical center, with data that would ultimately be publicly reported and factor into faculty physician incentive compensation. And we needed to do this with our existing EHR, enterprise data warehouse (EDW) and business intelligence (BI) reporting applications, without heavy investment in new technology. Clearly our registry development approach needed to change.

We thus took as our challenge designing and constructing a common informatics framework leveraging the existing EHR and EDW, which would be re-usable across multiple patient registries serving multiple specialty patient sub-populations.

### 1.2 Common Registry Framework

Must each EHR-based chronic disease registry and CQM project be treated as a "one-off" endeavor, with unique data capture methods and data extraction/reporting needs? Or can a common registry development framework be created? Certainly each condition covered by a registry has unique characteristics. And registries can be constructed for disparate primary purposes, such as for scientific or epidemiologic reasons, or for quality improvement [2]. So it's reasonable to assume no two reg-

istries will be the same. But despite these specific differences, some core similarities exist across chronic disease registries. Each is at the patient level, typically defined as patients with a given condition, or who've undergone a certain procedure or other treatment [2, 7]. Thus the categories of EHR data needed to include lists of patients are generally shared across registries: diagnoses/problems, procedures, results, medications [8].

Finally, the conceptual framework for calculating clinical quality measures (CQMs) is common across registries: certain patients are in the denominator, a subset of those are also in the numerator, and some are excluded. Thus a common framework becomes plausible for defining registries, and for capturing and retrieving core types of data from the EHR to calculate and store CQM measurements [9].

### 1.3 Structured Data Capture

One point of view holds that structured EHR data is too often incomplete or inaccurate to use for registries, clinical quality measures, and/or clinical research. Compared with highly-controlled data entry in typical clinical research, often via structured Case Report Forms, data enters an EHR during the less predictable process of actual care delivery. And in an enterprise EHR, many of the fields in a patient chart are shared across the entire network of providers and staff using the common EHR, and could be updated by any. Too few standards are promoted or adhered to about what should be entered on (or removed from) shared fields like the Problem List for instance, the argument goes; even with accepted standards, different physicians have differing mental models of the Problem List, leading to variation [10].

Despite this apparently bleak portrait, there's reason for optimism that an EHR can be a source of reliable data for clinical quality improvement projects and patient registries. One data quality truism is that "what gets used, gets better" (think of an employee's direct deposit bank account routing number in a Human Resources system – it's rarely wrong, and never stays incorrect for long). By this axiom, then the quality of clinical data fields actively

viewed and used in clinical care will progressively improve. And we've seen that happen with the Medication List in our EHR: initially often riddled with duplicate and inconsistent medication information, through constant use it's increasingly become the best source of medication information on our patients, and valuable to all clinicians involved in a patient's care.

A second reason for optimism is that only a relatively small number of EHR data elements are needed for registry population and CQM calculation. Analysis can then be focused on where these critical few data elements are best stored in the EHR, and how structured data entry/capture tools in the EHR can most optimally fit into existing clinical workflow and routines [11].

### 1.4 Dimensional Data Modeling

Dimensional data modeling consistently proves an effective data warehousing design approach. Compared with more complex data warehouse modeling approaches, dimensional modeling on average yields higher project success rates, produces data structures easy to understand and query, and supports iterative incremental development for delivering valuable new data content to data consumers early and often [12, 13, 14]. Dimensionally-modeled data warehouses take extracts of varied source transactional data extracts ("Facts") each at their fundamental "grain" (e.g. one row per patient encounter, or one row per medication order), and link them via master reference data ("Dimensions") to achieve analytic interoperability [12]. Many successful healthcare data warehouses employ a dimensional modeling approach, including governmental (national VA clinical data warehouse), EHR vendor-supplied enterprise data warehouses (Epic), and targeted business intelligence (BI) company solutions.

We wanted to explore how dimensional modeling could work both as a source for calculating electronic clinical quality measures (eCQMs) and for storing CQM results. A relatively modest number of grains of data from the EHR will contain the inputs needed for most eCQM numerator, denominator, and exclusion formulas. And

a single grain of data for CQMs can support the needs of multiple specialty patient registries, by recording data at the grain of one row per patient per measure per time period (and optionally per attributed provider). With a common source for calculation inputs and a common target for calculation output, calculating eCQMs using templated code becomes possible, streamlining ETL development.

### 1.5 Agile Registry Development as New Product Development

At first blush, developing a specialized chronic disease registry for quality improvement activities may seem far removed from the world of new product development and its association with turning a market opportunity into a compelling product made available for sale. The registry's clinical contents and value for scientific, epidemiologic and/or quality purposes take rightful focus, with the EHR and EDW tools a secondary consideration. A given clinical organizational unit or specialty society undertaking a registry might well be contemplating creating only a single registry, with no reason to consider a scalable framework for creating multiple registries.

Yet digging deeper, similarities emerge between new registry development and other kinds of new product development (NPD). By definition, NPD involves something not created before, as opposed to ongoing production. Unknowns initially abound. In one classification of projects into 4 zones labeled simple, complicated, complex, or chaotic, NPD lies in the "complex" zone [15]. Here solutions are not obvious but emerge through frequent inspection and adaptation, and innovation occurs. Developing a completely new specialty patient registry using EHR data falls squarely in this same zone.

In the complex zone where new product development occurs, agile project management (APM) consistently performs better than traditional project management (TPM). Key differentiating features of APM over TPM include use of iterative, incremental development during time-boxed "sprints", early and frequent production of working product features for rapid custom-

**Table 1** The present project's main points and ways to evaluate them.

Main point	How this project will establish/evaluate
A common framework can be constructed that is re-usable for multiple patient registries crossing multiple specialties and patient sub-populations	How we will establish this is correct: <ol style="list-style-type: none"> <li>1. Design and employ such an approach on a high-stakes real-world project on a tight timeline.</li> <li>2. Measure re-use: # and % of registries produced using the common registry framework, % of registries using templated ETL CQM calculation code, % of templated dashboards with reports and data visualizations. Compare with historical values for re-use (none).</li> </ol>
Structured data capture in the EHR – as soon after the information becomes known as possible – can be achievable, useful, and accurate	How we will establish this is correct: <ol style="list-style-type: none"> <li>1. Implement structured discrete data capture exclusively for all registries (no manual abstraction)</li> <li>2. Design an "audit"/discrepancy process wherever feasible (e.g. Problem List-based conditions vs. billing diagnosis-based conditions)</li> </ol>
How dimensional modeling can work for eCQMs	How we will demonstrate: <ol style="list-style-type: none"> <li>1. # of grains needed, # of grains re-used across registries (matrix)</li> <li>2. Ability to store all CQMs in a single dimensional data model</li> </ol>
Although registry creation and new product development at first seem different, they are actually similar; thus agile NPD methods will prove beneficial for registry creation	How we can empirically evaluate this claim: <ol style="list-style-type: none"> <li>1. Employ an agile methodology on a real-world EHR-based registry development project on a tight timeline</li> <li>2. Measure # of registries produced, # of specialties represented, # of patients actively followed, # of EHR tools constructed, cycle time for projects. Compare with historical approach values.</li> </ol>

er feedback, an obsession with testing included automated test-driven development, and a focus on a sustainable pace or "velocity" of delivering valued features to production in a regular rhythm. With APM, project success rates – while not 100% – are consistently higher than with TPM, customer satisfaction with the product is higher, a higher percentage of developed product features are actually used by customers, defects in production are lower, and value is realized from the product investment much earlier [16]. Benefits of agile methodologies were first unequivocally established in software-intensive product development, but have now been demonstrated in data warehousing / BI development projects [13, 14] and new product development in general. Given that new registry development is in effect NPD, the proven benefits of agile development methods are likely to translate well to use on new registry development [17].

### 1.6 The Present Project

The present project was intended to demonstrate and evaluate each of the 4 main points above (in sections 1.2–1.5) as shown in ► Table 1.

## 2. Objective

Our objective was to determine if the creation of EHR-based specialty registries for a large multi-specialty healthcare organization could be markedly accelerated by employing (a) a finite core set of EHR data collection principles and methods, (b) concurrent engineering of structured data extraction and data warehouse design using a common dimensional data model for all registries, and (c) agile development methods commonly employed in new product development.

**Table 2** Registry-defining information and optimal EHR location.

	Important Real-world Information	Optimal Location
1. Who are “my patients”?	Provider-patient relationships that span > 1 encounter	Care Team
2. Which have Condition X?	A given patient’s active health/medical issues	Problem List
3. Which have had exposure Y?	A given patient’s prior surgeries & major procedures, prior/current medications	Surgical History; Medication History

### 3. Methods

#### 3.1 Software

For this project we employed:

- our existing EHR clinical documentation, reporting, and population health modules (all from Epic Systems, Verona WI), and
- our existing enterprise data warehouse (EDW) and business intelligence (BI) tools: Microsoft SQL Server, SQL Server Reporting Services, SQL Server Analysis Services, Power BI (all from Microsoft Corporation, Redmond, WA).

#### 3.2 EHR-based Registries and Associated Data Collection Tools

Patient registries at the physician level: What’s needed? The most important real-world information needed to define most chronic disease registries [2] was matched

to the optimal location in the EHR for storing that information (► Table 2).

Data fields needed for clinical quality measure (CQM) calculations were derived sequentially by the following schematic:

Registry --> Measure --> Numerator/Denominator/Exclusion Formulas --> Data Fields

For each registry process and outcome measure, the numerator and denominator of the measure (and exclusions) were specified first in understandable sentence format, and then more precisely in mathematical formulas. Based on those mathematical formulas, the discrete data elements/fields needed to perform the measure calculation were then identified.

Each data element was assessed as being either “standard” EHR data, or a “custom” data element unique to the given registry. For each “custom” data element, the goal was to assess where in the workflow the in-

formation would first become knowable and to whom, and where data collection would fit most cleanly in the clinic’s workflow. Whenever possible one of 4 pre-defined EHR-specific types of structured data collection fields was selected for each custom data element (► Table 3).

As shown in ► Table 3, we defined “Standard Data” as information for which a standard structured location in the EHR already existed for use across our Health System – such as Date of Birth, Medication List, Problem List, Encounter Type, etc. Existing clinic operational workflows for populating this data were reinforced if needed.

We defined “Custom Data” as information needs unique to a given specialty registry, and for which new EHR tools and clinical workflows for data capture would need to be developed. To create a replicable framework, we chose 4 types of extractable fields that could fit for the vast majority of clinical workflows. Nurses and Medical Office Assistants (MOAs) document extensively in EHR Flowsheets, and adding custom registry-specific flowsheet tools typically fit best into their workflow. Providers at our institution are more likely to interact with Forms which can be presented selectively into their workflow based on sophisticated rules, either non-interruptively or interruptively, and responded to with rapid selection from presented options (typically as buttons). These forms populate extractable Custom Data Elements. Both providers and RNs/MOAs can potentially enter Results into the chart – where a desired data element was essentially a Result, a “Results Console” could be provided within the normal office visit documentation workflow for entering this registry-specific data. Finally, for patient-reported outcomes, we used the EHR’s patient questionnaire framework; these custom questions could be delivered to the patient via the patient portal, or via a tablet or on-screen during their clinic visit.

General use cases for registry data collection are shown in ► Online Appendix A. Often a specialty clinic employed a combination of EHR tools for registry data collection – these details could be captured on a use case diagram specific for that registry. Where a registry employed a specific se-

**Table 3** EHR sources for data elements needed.

Type of Data	Examples	Discrete Data Fields
Standard Data: Already captured in EHR in standard locations (or could be)	Age, Medications, Diagnoses, Surgical History, Insurance Coverage, etc.	Use EHR-standard fields
Custom condition-specific Data: • Captured by Clinical Staff	Medical reason for exclusion from registry; Clinician assessment of response to treatment	1. Flowsheets (used by RNs/MOAs) 2. Forms with Custom Data Elements (used by MDs/APPs in our institution) 3. (Custom Test) Results (populated by EHR tool designed for rapid numeric result entry)
Custom condition-specific Data: • Captured by Patient	Patient-reported outcomes	4. Patient-entered Questionnaires (via EHR’s patient portal or in-clinic EHR tool)

Fact Grain (one row per:)	Dimensions (analyze by:)									
	Date	Provider	Patient	Encounter/ Enc Type	Clinical Org Unit	Location	Diagnosis	Treatment/ Procedure	Medi- cation	Quality Measure
Encounter	X	X	X	X	X	X	X	X		
Order (Test/Procedure)	X	X	X	X	X	X	(X)	X		
Medication Order	X	X	X	X	X	X	(X)		X	
Problem	X	X	X	X	X	X	X			
OR Procedure	X	X	X	X	X	X	X	X		
Results	X	X	X	X	X	X	X	X		
Patient-entered Questionnaire Answers	X	X	X	X	X					
Flowsheet Entry	X		X	X	X	X				
Custom Data Concept	X		X							
Clinical Decision Support Alert Display	X	X	X	X	X	X				
Clinical Quality Measure Value	X	X	X							X

Figure 1 Dimensional data model (conceptual level).

quence of data collection activities, a UML activity diagram (swimlane workflow diagram) depicted the sequence and role responsible for each step (► Online Appendix B).

### 3.3 Data Extraction and Data Warehousing

By deciding to use standard EHR fields and a core set of EHR structures for capturing custom data, we could then in parallel develop a replicable extract-transform-load (ETL) process from the EHR into the Enterprise Data Warehouse (EDW). The existing dimensional data model for the EDW employed standard dimensional modeling techniques [15], and already included Patient, Encounter, Order, Result, Medication, Problem/Diagnosis, OR (Surgical) Procedure, and Professional Charges data, among others. The EDW dimensional data model was extended to include new grains of data for each of the core four custom data structures in ► Table 3 (see ► Figure 1). The EHR conceptual data schema (► Online Appendix C) was mapped to the target EDW dimensional model (► Online Appendix D). Construction of ETL of all EHR data of these types then proceeded in parallel with specialty registry development. One exception was flowsheet

data, as the amount of this data type in the EHR was massive – here the ETL framework was built to readily extract data for specific flowsheet template ID numbers used in the registry project simply by updating a table of included template IDs. Additional grains of data for standard EHR data were added (and associated ETL constructed) when found to be necessary for registry measure calculation. Again, 100 percent of the EHR data of each grain then was extracted into the EDW, irrespective of whether known to be associated with a registry project, so that any future project needing the same grain of data could also be supported.

### 3.4 Clinical Quality Measures

For each of a registry's eCQMs, starting from the overall Registry population, formulas for the denominator, numerator, and any exclusion sub-populations were defined – first in sentence form, then in mathematical formulas with reference to specific data elements (see ► Online Appendix E for Venn diagram of populations). A target dimensional data model was designed with a single Fact table to hold all eCQMs from all registries, at the grain of one row per patient per time period per measure per attributed provider.

The corresponding Dimension tables then were Patient, Time Period, Quality Measure and Provider (► Online Appendix F). Although the original conceptual model envisioned assigning provider attribution by Care Team (to reinforce optimal use of this field in the EHR), to ensure the lists viewed by providers were relevant for quality improvement actions, attribution logic was also added to exclude patients who had not been seen within the past 24 months by the provider (► Online Appendix G).

Making use of (a) the common set of source Fact tables in the EDW dimensional data model and (b) the common target dimensional data model (single Quality Measure Fact table), SQL code could be written to execute the mathematical formulas for calculating the numerator, denominator, and exclusion status for each patient-measure-time period combination. Stored procedures were used to execute the SQL code. One stored procedure was assigned for each eCQM, with responsibility for assigning the Numerator, Denominator, and Exclusion status for each patient and populating the common Quality Measure fact table. The similarities in source and target tables across measures enabled construction of a single stored procedure template. This template was re-used as

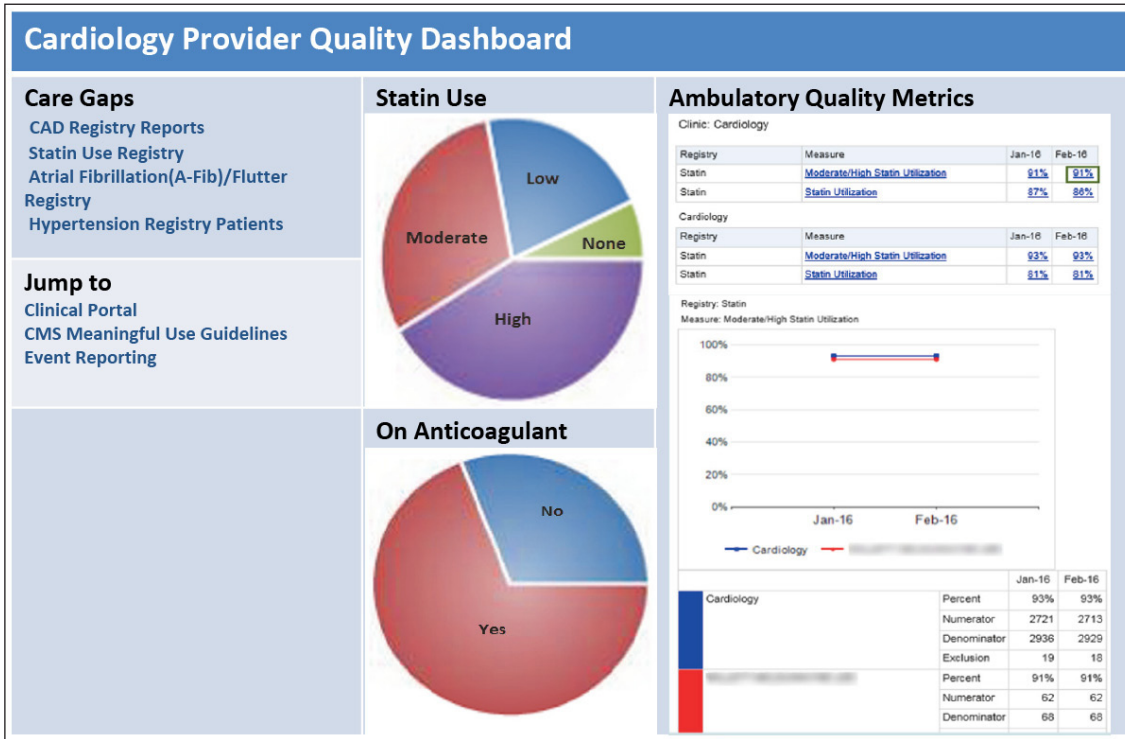


Figure 2 Registry dashboard.

the starting point for constructing each eCQM-specific stored procedure. Additionally, the stored procedures' references to specific EHR items (e.g. diagnosis or procedure grouper IDs) were stored in tables, so that the ETL code housed inside the stored procedures could be table-driven. If an EHR grouper item changed, the ETL code could be correspondingly updated simply by updating the reference table, without having to re-write the ETL stored procedure code.

### 3.5 Registry Data Reporting/ Visualization Tools

Since the target audience for the specialty registries primarily consisted of clinicians and clinical managers, registry reporting was done through the EHR in which those persons primarily did their daily work. One quality-specific dashboard was created in the EHR for each specialty. If a specialty had more than one registry, their single dashboard included information from all registries. Each specialty dashboard included both native EHR information (real-time or near-real-time), and performance measures (eCQMs) calculated in the EDW weekly (► Figure 2).

Native EHR information included graphical displays of registry patient characteristics, and registry-specific patient lists. Patient lists displayed columns of demographic and clinical care gap information (► Online Appendix H), as well as a detailed report about any single selected patient. Using native EHR functionality, from these lists a clinician could directly enter the full medical record of a selected patient. Additionally, the list could be filtered, e.g. to show only patients with a specific care gap (see ► Online Appendix A for general Reporting use cases). Bulk ordering or bulk communications could then be done for the selected group of patients to close care gaps. Bulk communications respect the communication preferences of the patient (electronically through the patient portal or by regular mail). Outreach via telephone to selected patients could also be documented, and follow-up at specific time intervals scheduled.

Each specialty's dashboard also displayed vetted eCQM performance measure data sourced from the EDW's Quality Measure Fact and Dimension tables (► Online Appendix I). The logged-on user's information was passed to the EDW via a web-based call, so that provider-spe-

cific information could be returned. For an individual provider, each measure was graphed to show the performance trend over time for both the provider themselves and their specialty as a whole; the exact values were also provided in table form. Drill-down capability was provided from this graph to a list of all patients making up a given data point, with the Numerator, Denominator, and Exclusion status for each patient (presented as 0 or 1).

Because a single data model holds all CQMs, additional report views could be constructed enabling interactive exploration of CQM trend(s) across any selection of specialty registries (► Online Appendix J).

**Secondary uses of data:** Use of EHR data in registries for internal quality improvement activities is considered Institutional Review Board (IRB)-exempt at our institution. Secondary use of registry data for research purposes requires IRB approval. External release of data outside the institution for secondary use (data-sharing with another organization, not submission of a research article for publication) requires a signed data-sharing agreement, even for non-personally identifiable data. External release of personally-identifiable

data additionally requires signing of a Business Associates Agreement (BAA) with the party to whom the data is being released, including provisions restricting further secondary disclosure, to conform with privacy regulations in the United States.

### 3.6 Agile Project Management

Even with the repeatable framework designed above, delivering new registries across multiple specialties in a short time window posed challenges. Additionally, refinement of registry build after initial real-world use by clinicians was expected. The number of developers available to work on registry tool construction at any one time was finite. Accordingly, we turned to agile project management (APM) techniques to maximize the amount of value we could create in the time available [16].

From the start, a co-development mindset was adopted, with representatives of four key groups participating in all phases of the project: clinic operations, quality and performance improvement, EHR developers, and business intelligence (BI) developers. This aided greatly in shared communication and in creating shared mental models of the problems being addressed and the solution designs.

Requirements documentation was kept lightweight. A high-level project charter for each specialty registry was co-drafted in advance between the quality department and the clinical specialists, outlining the clinical focus of the registry, the outcomes being sought, any process measures desired to prove quality care delivery, and an initial concept of the types of data collection desired (e.g. any specific patient-reported outcome questionnaires). “User stories” were written for the overall registry, in the standard form of: “As a <person’s role>, I want <some new capability>, so that <a benefit will be achieved>.” Simple to write and to read, this technique is widely used in the agile community for rapidly clarifying among all parties the “who?”, “what?” and “why?” of a requested new product or capability [16, 18]. Acceptance criteria were written for each user story as simple bulleted lists or tables of desired outcomes, and provided further definition of “what success looks like” on any given registry

project. Where useful, more detailed user stories and acceptance criteria were written for individual key components of a registry, such as a clinical decision support alert. Agile modeling [19] and model-driven development [20] were employed judiciously to help create shared “mental models” of complex registries with multiple components (►Online Appendix K, ►Online Appendix L).

Time-boxed iterative development formed the cornerstone of the agile project management (APM) process employed. Two-week iterations were scheduled for the duration of the project. Each registry was initially allotted 2 to 3 two-week iterations for building of the EHR data collection and clinical decision support components. An overall project plan laid out which registries would kick off at the start of each two-week iteration – typically this was 3 registries, leading to a staggered development timeline that controlled the amount of work-in-progress (WIP) at any one time to fit within the capacity of the available developer teams (► Online Appendix M).

The goal of each 2-week development was production of working product within the EHR. That is, the development lifecycle of detailed design, build, and test was all to be accomplished within a single 2-week iteration. To do this, each large registry project was “sliced” into deliverable EHR components that could be accomplished within a 2-week iteration. Demos of newly developed working product were held at the end of each 2-week iteration. Although components were developed production-ready in 2-week iterations, the build was held in a Test (pre-production) environment until all components of a given registry were ready to be delivered to the specialty clinicians as a package, and any training and instructional materials had been delivered. The entire registry set of EHR tools and workflows was then released to Production on a specific release (“go-live”) date for each registry. Not shown in ►Online Appendix M are the iterations for incrementally building the data warehousing, analytic, and reporting features for the registries, which proceeded in parallel, and produced a common framework shared by all registries.

## 4. Results

### 4.1 Primary Quantitative Results

The primary goal of implementing this common EHR-EDW framework was to determine if the process of creating EHR-based specialty registries could be markedly accelerated. Prior to this project, we had implemented 2 EHR-based registries over a 2-year period, both for primary care conditions, and neither of which achieved incorporation into normal clinical quality monitoring or reporting. Following the methods described, in calendar year 2015, a marked increase in delivery of EHR-based registries into our production EHR and active clinical care was accomplished (► Table 4). All registries were created using only discrete EHR data capture, and all constructed using the common ETL and EDW dimensional model, and common reporting mechanism.

A total of 43 specialty registries were constructed for 30 specialties. Each specialty was delivered a specialty-specific EHR quality dashboard for use in visualizing their registries’ data and accessing actionable patient lists. During this phase of the project, a total of 111 EHR-based data collection tools were created. 163 measures were also created, all derived from EHR data without need for manual abstraction. The eCQMs included both process measures and patient-reported outcomes measures, with completion of >2,100 patient-reported questionnaires during the initial project period.

To accomplish this scope of registry and eCQM development in one year required a re-usable data collection and data analytic framework. Thus we aimed to establish a data capture “pipeline” from EHR data collection tools into the EDW, with extensive re-use of EDW tables across registries. Re-use of major dimensional model grains of data is shown in ►Figure 3: rows shaded in orange indicate the four types of data employed for all custom data capture fields.

Employing a common EHR and EDW framework and an agile methodology substantially accelerated EHR-based specialty registry development from our baseline of 52 weeks to 14 weeks. These registries quickly provided the ability to actively manage and measure care for greater than 60,000 patients.

**Table 4** Calendar year 2015 registry development production and patient enrollment.

# of registries constructed	43
% registries built using common EHR/EDW framework for EHR data capture, ETL, and EDW dimensional data storage (n=43)	100%
% registries using common code template as basis for code calculating the registry's CQMs from dimensional data (n=43)	100%
# of specialties provided >=1 registry	30
% specialties provided dashboard with EHR and EDW data, built using common framework (n=30)	100%
# of EHR-based data collection and CDS tools created	111
# of eCQMs created	163
# of two-week iterations per registry for EHR feature development (average)	4 iterations (8 weeks)
Average total duration of initial registry project per specialty	14 weeks
# of patients actively managed on one or more registries	>60,000
# patient-reported outcome questionnaires completed	>2,100

## 4.2 Additional Qualitative Results

Beyond these numeric results addressing the primary objective of this project, we observed additional secondary qualitative and subjective valuable outcomes: team skill growth and clinician acceptance of registries. These are described further in ►Online Appendix N.

## 5. Discussion

Constructing even a single EHR-based specialty registry with associated EHR data collection and clinical decision support tools can seem daunting and time-consuming, let alone developing subsequent data extraction, database storage, and interactive registry reporting. However, we proposed that:

- a common framework can be constructed which is re-usable across multiple patient registries, serving multiple specialties and patient sub-populations
- data capture in the EHR for registry definition and eCQM calculation can prove clinically practical and analytically useful
- dimensional modeling can work for eCQM derivation and storage, and
- agile methodologies will work well for EHR-based Registry development, just

as they do for new product development in general.

Our central finding was that employing just such a shared framework with agile methods markedly shortened specialty registry development time. During 2015, 43 registries using only EHR structured data were constructed in an average of 14 weeks each (8 weeks EHR development time), leading to management of >60,000 registry patients and with >2,000 patient-reported outcomes. Additional registries and quality measures continue to be added using the same framework.

### 5.1 Designing in Data Quality for Structured EHR Data Capture

Given this work, we believe that EHR data can and should be used both for registry definition and for eCQM calculation, by designing in data quality from the start. Despite objections that structured EHR data is too-often incomplete or incorrect, a finite – and small – number of data elements are actually used for registry purposes.

For these crucial few data elements, validity can be enhanced by (a) explicitly deciding the best place in the patient's electronic health record to store a given piece of real-world information (e.g., current medications), and (b) assessing where in

the clinical workflow this information first becomes known, and who is best positioned to enter it naturally as part of their clinical activities (patient, medical office assistant, nurse, or physician). Armed with this information, an optimal EHR-based data collection tool can be selected and constructed. Additionally helpful is making these critical few data elements highly visible in normal clinical workflows so that data quality (completeness, accuracy, internal consistency) can be readily seen, and corrected if need be. Clinicians sponsoring quality improvement registry projects want accurate patient and CQM data. Focusing on the critical few EHR fields aligns incentives to complete and validate them.

A “halo effect” can ensue, as clarifying and optimizing clinical data crucial for registry inclusion and eCQM calculation yields follow-on benefits. Almost by definition, these select structured data elements provide important clinical context, useful both clinically and analytically [21].

### 5.2 Dimensional Data Modeling Holds Promise for Analytic Interoperability

In database design, the closer a data model reflects the real world, the more robustly it can handle future use cases of all types. This applies to data warehouse design as well, where the use cases are analytic. Dimensional data modeling strives to match real world processes and transactions, along with the analytically-useful context surrounding those transactions.

Disparate chronic disease registries often share need for information drawn from the same clinical processes and transactions, with associated measures. In dimensional jargon, these are “Facts”: natural data grains of one row per Encounter, per Order, per Test Result, per Medication Prescribed, per Diagnosis/Problem assigned, etc. And a finite set of core Dimensions provides a robust ability to “slice and dice” this Fact data: by Date, by Patient, by Encounter (Type), by Location, by Diagnosis, by Procedure, etc. (►Figure 1, ►Online Appendix J). With a matrix consisting of grains of data on rows, and context dimensions on columns, potential linkages among disparate data grains become



<u>Data Warehouse Table</u>	<u># Registries Using</u>	<u>Fact Table Grain. One row per:</u>
FACT_Encounter	26	Patient appointment/encounter
FACT_PatientProblemGroupier	15	Patient problem in diagnosis grouper
FACT_Flowsheet	14	Flowsheet cell entry
FACT_CustomDataValue	14	Custom data field entry
FACT_Questionnaire	12	Question answer
FACT_Alert	8	CDS Alert seen by clinician
FACT_OrderResult	7	Test result
FACT_PatientProblem	7	Patient problem
FACT_AllergenGroupierOrder	6	Patient allergy in allergen grouper
FACT_MedicationGroupierOrder	6	Medication order in medication grouper
FACT_ProcedureGroupierOrder	6	Procedure order in procedure grouper
FACT_ImmunizationGroupierOrder	3	Patient immunization in grouper
FACT_ProcedureOrder	2	Procedure order placed

**Figure 3**  
Re-use of data grains across specialty registries.

clearly visible – providing analytic interoperability even if the source systems are not integrated.

These same concepts can be extended to analytic interoperability not only for linking data from different specialties, but also for linking data from different EHRs and different organizations. To accomplish this, the master reference Dimensions would need to be shared (or “conformed”) across sites. That’s solvable where widely accepted standards are in place (e.g. LOINC for lab results, SNOMED for diagnoses, etc.), by using standards-based code sets or groupers (► Online Appendix O). Master patient and provider data would be needed as well. Dimensional modeling thus provides one possible technical approach to achieving the data linkage needed to accelerate the path to a Learning Health System [22].

### 5.3 Agile Methods Work Well for EHR-based Registry Development

Recognizing the similarities between EHR-based registry development and new product development opens the door to employing agile principles and practices, and harvesting their proven benefits. Agile practices especially helpful for new EHR-based registries include:

- Co-development (e.g. by clinical, quality, EMR, and data warehousing/analytics team members)
- Light-weight project artifacts, such as User Stories, clear acceptance criteria, and agile modeling

- Time-boxed, iterative and incremental delivery of working features (e.g. every 2 weeks)
- Frequent demos to clinical customers of working product, to promote a rapid feedback cycle.

Benefits realized by this approach include early risk reduction, early return on investment by delivering working and valued features early, minimizing waste spent building features ultimately never used, defect reduction through short feedback loops, and a sustainable pace for developers.

Thus by adopting agile methods over traditional waterfall development methods, if you are sponsoring a registry, you will get more functioning registry capabilities for any given resource investment. If developing a registry, you’ll be focused on creating valued features that are actually used, while working at a sustainable pace. And if wanting to use registry data, you’ll receive actionable data much earlier that you can begin employing in your efforts to improve care for patients.

### 5.4 Issues Encountered

Despite active clinician engagement and opportunities to refine EHR build based on short feedback cycles, a few registries still proved little-used initially after going live in the production environment. Potential reasons observed include:

- Some registries required an extra manual step to select which patients to “enroll” in a registry, rather than basing on

a diagnosis or performed procedure. Thus enrollment relied more heavily on motivation of busy clinic staff, with no directly visible reinforcing benefit.

- Varying from our principle to construct patient-level registries. Specifically, in one case we attempted a biopsy-level registry which proved difficult to accomplish, and we were unable to leverage patient-level EHR constructs already confirmed to work.
- Communicating and marketing the registry concept from the initial requestor/physician champion to other physicians and/or clinic staff did not always occur effectively.

To address the first two contributing causes, we reaffirmed our guiding principle that each registry be based on patients with either (a) a common condition, or (b) a common exposure (to a procedure or medication). We also reaffirmed a focus solely on patient-level registries for now – distinguishing between registries of patients who’ve undergone a procedure (one row per patient) vs. a registry of procedures (one row per procedure performed).

Regarding effective transmission of registry adoption across all clinicians in a specialty, we appreciate the challenging position we ask our registry physician champions to fill. In discussing quality improvement registries, the AHRQ User’s Guide for registries calls out this crucial role:

“Yet, the common theme for both local and national QI registries is that the local champions must be successful in actively engaging their colleagues in order for the program to go beyond an ‘early adopter’ stage and be sustainable within any local organization. Once a registry matures, other incentives may drive participation (e.g., recognition, competition, financial rewards, regulatory requirements), but the role of the champion in early phases cannot be overstated.” [2: Vol 2, p. 172]

Accordingly, during new registry development we have begun including discussions early on how best to communicate and “market” the registry to all clinicians within that specialty.

## 6. Conclusions

A new EHR-based registry project often proves to be a complex and lengthy endeavor, involving sequential creation first of data collection tools and clinical decision support tools in the EHR, followed by creation of registry-specific data extractions from the EHR for analysis. The next registry request then starts from scratch. Such an approach will have difficulty scaling up to meet the increasing demand for multiple specialized disease registries needed to support value-based care and population health. We set out to determine if the creation of EHR-based specialty registries could be markedly accelerated by employing a common framework of EHR and EDW data structures for all registries, and by applying agile development methods. Using this approach, in calendar year 2015 we developed a total of 43 specialty chronic disease registries, with 111 new EHR data collection and clinical decision support tools, 163 new clinical quality measures, and 30 clinic-specific dashboards reporting on both real-time patient care gaps and CQM measure performance trends. This study suggests concurrent design of EHR data collection tools and reporting can quickly yield useful EHR structured data for chronic disease registries, and bodes well for efforts to migrate away from manual abstraction. This work also supports the view that new EHR configuration

shares much in common with new product development, and consequently adopting agile principles and practices can yield similar benefits by delivering valued, high-quality EHR features early and often.

### Acknowledgment

The authors thank the following UT Southwestern leaders: Dr. Daniel Podolsky, Dr. Bruce Meyer, Kirk Kirksey, and Suresh Gunasekaran, for contributing ideas and resources to the ambulatory quality registry development effort. They also extend deep appreciation to Margaret Parks for invaluable assistance with the preparation of this article, and to their teammates and colleagues at UT Southwestern who brought so much creativity and energy to this project.

## References

- Porter, ME. What is Value in Health Care? *N Engl J Med* 2010; 363(26): 2477–2481.
- Gliklich RE, Dreyer NA. *Registries for Evaluating Patient Outcomes: A User's Guide*. 3rd ed. Rockville, MD: Agency for Healthcare Research and Quality; 2014.
- Blumenthal D, Chernof B, Fulmer T, Lumpkin J, Selberg J. Caring for High-Need, High-Cost Patients – An Urgent Priority. *N Engl J Med* 2016; 375 (10): 909–911.
- Pruitt S, Annandale S, Epping-Jordan J, Fernandez Diaz JM, Khan M, Kisa A, Klapow J, Solinis RN, Reddy S, Wagner E. *Innovative Care for Chronic Conditions: Building Blocks for Action*. Geneva: World Health Organization; 2002.
- Larsson S, Lawyer P, Garellick G, Lindahl B, Lundstrom M. Use Of 13 Disease Registries In 5 Countries Demonstrates The Potential To Use Outcome

## Abbreviations and Acronyms

AHRQ: Agency for Healthcare Research and Quality (U.S.A.); APM: agile project management; APP: advanced practice provider [e.g. nurse practitioner, physician assistant]; BI: business intelligence; CQM: clinical quality measure; eCQM: electronic clinical quality measure; EDW: enterprise data warehouse; EHR: electronic health record; EMR: electronic medical record; ETL: extract-transform-load [data]; MOA: medical office assistant; OR: operating room; QI: quality improvement; TPM: traditional project management; UML: Unified Modeling Language

- Data To Improve Health Care's Value. *Health Affairs* 2011; 31(1): 220–227.
- Backus LI, Gavrilov S, Loomis TP, Halloran JP, Phillips BR, Belperio PS, et al. Clinical Case Registries: Simultaneous Local and National Disease Registries for Population Quality Management. *JAMIA* 2009; 16(6): 775–783.
  - Emilsson L, Lindahl B, Köster M, Lambe M, Ludvigsson JF. Review of 103 Swedish Healthcare Quality Registries. *J Intern Med* 2014; 277(1): 94–136.
  - Wright A, McGlinchey EA, Poon EG, Jenter CA, Bates DW, Simon SR. Ability to Generate Patient Registries Among Practices With and Without Electronic Health Records. *Journal of Medical Internet Research* 2009; 11(3): e31.
  - Benkert R, Dennehy P, White J, Hamilton A, Tanner C, Pohl J. Diabetes and Hypertension Quality Measurement in Four Safety-Net Sites. *Appl Clin Inform* 2014; 5(3): 757–772.
  - Krauss JC, Boonstra PS, Vantsevich AV, Friedman CP. Is the problem list in the eye of the beholder? An exploration of consistency across physicians. *JAMIA* 2016; 23(5): 859–865.
  - Cusack CM, Hripscak G, Bloomrosen M, Rosenbloom ST, Weaver CA, Wright A, et al. The future state of clinical data capture and documentation: a report from AMIA's 2011 Policy Meeting. *JAMIA* 2013; 20(1): 134–140.
  - Kimball R, Ross M. *The Data Warehouse Toolkit: the Definitive Guide to Dimensional Modeling*. Indianapolis: Wiley; 2013.
  - Collier K. *Agile Analytics: A Value-Driven Approach to Business Intelligence and Data Warehousing*. Pearson; 2012.
  - Hughes R. *Agile Data Warehousing Project Management: Business Intelligence Systems Using Scrum*. Elsevier (Morgan Kaufmann); 2013.
  - Rubin KS. *Essential Scrum: a Practical Guide to the Most Popular Agile Process*. Upper Saddle River, NJ: Addison-Wesley; 2012.
  - Larman C. *Agile and Iterative Development: A Manager's Guide*. Upper Saddle River, NJ: Addison-Wesley; 2004.
  - Greenberg AE, Hays H, Castel AD, Subramanian T, Happ LP, Jaurretche M, et al. Development of a large urban longitudinal HIV clinical cohort using a web-based platform to merge electronically and manually abstracted data from disparate medical record systems: technical challenges and innovative solutions. *JAMIA* 2016; 23(3): 635–643.
  - Cohn M. *User Stories Applied: For Agile Software Development*. Upper Saddle River, NJ: Addison-Wesley; 2004.
  - Ambler SW. *Agile modeling: effective practices for eXtreme programming and the unified process*. New York: Wiley; 2002.
  - Rosenberg D, Stephens M. *Use Case Driven Object Modeling with UML: Theory and Practice*. New York: Apress; 2013.
  - Wright A, McCoy AB, Hickman T-TT, Hilaire DS, Borbolla D, Bowes WA, et al. Problem list completeness in electronic health records: A multi-site study and assessment of success factors. *International Journal of Medical Informatics* 2015; 84(10): 784–790.
  - Ainsworth J, Buchan I. Combining Health Data Uses to Ignite Health System Learning. *Methods Inf Med* 2015; 54(6): 479–487.