5-17-2021

Ranking Methodology to Evaluate the Severity of a Quality Gap Using a National EHR Database.

Shivani Sivasankar

An-Lin Cheng

Mark A. Hoffman

Ranking Methodology to Evaluate the Severity of a Quality Gap Using a National EHR Database

Shivani Sivasankar, M.S.,[1,2] An-Lin Cheng, Ph.D.,[1] and Mark Hoffman, Ph.D. [1,2]

[1]University of Missouri-Kansas City School of Medicine, MO; [2]Children's Mercy Hospital, Kansas City, MO

Abstract

Selecting quality improvement projects can often be a reactive process. In order to demonstrate a data-driven strategy, we used multi-site, de-identified electronic health record (EHR) data to prioritize the severity of a quality concern: inappropriate A1c test orders for sickle cell disease patients in two randomly chosen facilities (Facility A & B). The best linear unbiased predictions (BLUP) generated from Generalized Linear Mixed Model (GLMM) was estimated for all 393 facilities with 37,151 SCD patients in the Cerner Health Facts™ (HF) data warehouse based on the ratio of inappropriate A1c orders. Ranking the BLUP after applying the GLMM indicates that the facility A being in the second quartile may not have a quality gap as significant as facility B in the top quartile for this quality concern. This study illustrates the utility of multisite EHR data for evaluating QI projects and the utility of GLMM to enable this analysis.

Introduction

One of the five competencies defined by the Institute of Medicine as essential for all healthcare professionals is the ability to apply principles of Quality Improvement (QI) to evaluate and improve systems performance [1]. The healthcare provider is expected to select appropriate quality indicators and maintain a focus on the areas considered most likely to improve patient care and clinical outcomes [2]. National standards, such as the Centers for Medicare and Medicaid Services (CMS) Core Measures and the National Quality Forum (NQF), provide important guidance for high level quality metrics [3,4]. Often QI initiatives are launched in response to local concerns including error prevention initiatives or efforts to improve time-based metrics [5]. It is not uncommon for the number of potential QI projects to exceed the capacity to complete the work, requiring leaders to prioritize among multiple options. By determining the severity of a quality gap compared to peer organizations, healthcare organizations can strategically prioritize QI projects, focus on the greatest opportunities and allocate time and resources to the more complex, yet critical, qualitative aspects of transforming care to improve population health with patient-centered affordable care [6].

Given the complexity of health care, assessing a quality gap is a dynamic and challenging process. Data is widely available for mandated metrics, enabling organizations to determine whether they are aligned with national averages or deviate substantially from the mean. For example, the National Quality Strategy developed by the NQF provides guidance on prioritizing performance measure gaps for adult immunizations [7]. The CMS has defined core sets of quality measures in specific clinical areas, including cardiology, gastroenterology, HIV and hepatitis C, medical oncology, obstetrics and gynecology, orthopedics, Pediatric Accountable Care Organizations (ACOs), Patient Centered Medical Homes (PCMH), and primary care [8]. However, high quality data to support prioritization of potential QI projects for many other clinical processes are not as widely available.

Other than attaining nationally mandated quality metrics, most quality improvement initiatives are based on local concerns and can often be reactive to a real or perceived crisis [9–11]. Contextualizing the severity of a quality concern can be difficult in the absence of data from peer institutions. Electronic Health Record (EHR) systems can advance the quality of care by providing access to patient health information, monitoring compliance with standard measures of quality so that patients receive guideline recommended care and reduce medical errors through decision support [12]. Individual organizations may use a local data warehouse populated by EHR data to evaluate the quality concern based on previous performance within that institution, for example when often quality improvement (QI) projects are selected in reaction to a recent adverse event [13,14]. While this is frequently useful, local data warehouses lack the broader data necessary to contextualize the severity of a quality concern. This limitation may result in failure to recognize deeper systemic quality gaps in which an organization varies more significantly compared to their peers. Large databases of aggregate, de-identified, data from healthcare providers throughout the U.S. can be useful to compare measures of outcome and practice across several facilities.

Multi-institutional data warehouses (MDW) are distinguished by the consolidation of disparate EHR data sources across several independent, non-affiliated organizations and their use of data structures that are optimized for

queries[15]. Health care facilities can achieve improvement in patient outcomes by participating in MDW [16]. Utilization of a MDW is an effective path to improvement as it encompasses a multifaceted strategy, the ultimate target of which is to decrease the uncertainty inherent to the process. Some of the commonly used aggregate databases include the Nationwide Inpatient Sample (NIS), the Kids' Inpatient Database (KID), the Nationwide Emergency Department Sample (NEDS), the Pediatric Health Information System (PHIS) and the American College of Surgeons National Surgical Quality Improvement Program (ACS NSQIP) [17–21]. These databases contain detailed, longitudinal analysis of hospital encounters and serve as a valuable resource for research and QI. Another such data resource, Cerner Health Facts (HF), has been demonstrated to have frequency of diagnoses codes consistent with the HCUP National Inpatient Survey [22], and has been found useful to evaluate national trends of several health care related research questions [23–26]. HF is distinguished by deeper granularity than the other examples. One challenge in using MDW is accounting for varying characteristics of the contributing facilities. HF includes large and small providers, urban and rural facilities, public and private organizations and geographically diverse contributing sites.

When patients are nested in clusters such as facilities, observations within the same cluster are likely to be correlated. Generalized linear mixed models (GLMM) including both fixed and random effects have been proposed to analyze correlated data which could adjust for facility level variation. Best Linear Unbiased Prediction (BLUP) estimated from the model after adjusting the bias can be used as an index to predict the severity of QI concern. As a proof of concept, in this study we use this methodology to evaluate and rank one of the global quality concerns in health care, inappropriate use of A1c orders for sickle cell disease patients. The glycated hemoglobin test (A1c) test to assist in the diagnoses and management of diabetes may result in false results for sickle cell disease patients due to the abnormal hemoglobin structure and shorter lifespan of their red blood cells [27–31]. Professional organizations such as the American Diabetes Association, National Institute of Diabetes and Digestive and Kidney Diseases and the National Glycohemoglobin Standardization Program recommend the use of alternative tests instead of A1c for patients with SCD [32–34]. We describe a novel application of a multi-institutional EHR data warehouse by resolving the requirement to handle variations between facilities in order to contextualize the severity of inappropriate A1c orders in sickle-cell patients by comparing two random facilities among its peers. We also determine the extent to which facility characteristics contribute to this quality concern.

Materials and Methods

Data source

This study used the de-identified Health Facts data warehouse (Cerner Corporation, Kansas City, MO), which contains longitudinal patient data systematically extracted from the EHR at participating institutions and includes encounter data (emergency, outpatient, and inpatient), patient demographics (age, sex, and race), diagnoses and procedures, laboratory data and facility characteristics (census region, number of beds, acute setting and teaching versus nonteaching status). The release used for this work (2016) consists of 386.2 million encounters and 4.3 billion lab results from 64 million patients in 863 healthcare facilities. All admissions, in patient medication orders and dispensing, laboratory orders and specimens are date and time stamped, providing a temporal relationship between treatment patterns and clinical information. All data are de-identified in compliance with the Health Insurance Portability and Accountability Act (HIPAA) before being provided to the investigators. Longitudinal relationships between patient encounters within the same health system are preserved. The University of Missouri Kansas City (UMKC) Institutional Review Board has determined that work with Health Facts is considered non-human subjects research.

Extraction of study cohorts from database

All patients with a sickle cell disease (SCD) diagnosis were included in the study cohort based on the International Classification of Diseases, Ninth Revision and Tenth Revision, Clinical Modification codes (ICD-9-CM and ICD-10-CM). These codes were selected based on clinical judgement and the criteria specified in the Phenotype Knowledgebase (PheKB) resource which identifies patients with confirmed SCD diagnosis with a positive predictive value of 99.4% and a sensitivity of 99.4% using ICD-9-CM diagnosis codes [35]. The resulting definition groups (from ICD-9-CM codes) were combined with ICD-10-CM codes to identify the sickle cell disease patient cohort. Encounters before 2011 and after 2016 were excluded from the analysis because Health Facts data architecture was updated in 2008-2009, and the data for 2017 was incomplete. This cohort was evaluated for the presence of A1c orders placed after the first sickle cell diagnosis. Laboratory tests in Health Facts are associated with Logical Observation Identifier Names and Codes (LOINC) values. The LOINC codes which were used to indicate A1c orders were 55454-3, 41995-2, 4548-4, 17855-8, 4549-2, 17856-6. This SCD patient cohort was divided into two groups based on the

presence or absence of A1c encounters: SCD patients with at least 1 A1c encounter and SCD patients without any A1c encounters.

Model building

Two facilities (Facility A and facility B) were randomly chosen to observe the difference in their ranking before and after applying the GLMM. In order to contextualize QI concerns, we described our measures as rates, with the numerator indicating how many times the measure has been met and the denominator indicating the opportunities to meet the measure. The outcome measure for this analysis is the ratio of SCD patients with A1c test orders over the total number of SCD patients in the facility. Contextualizing the two random facilities relative to the other facilities in HF posed two challenges. First, the ranking process needed to be adjusted to address the repeated measures of the facility and facility level differences as each facility can vary from small ambulatory clinics to large hospitals with more than 500 beds. Second, the variation in the denominator (total number of SCD patients) between the facilities need to be accounted for, as each facility encompasses differing SCD patient populations. A single level logistic regression model which predicts outcome from the facility level predictor was extended to multilevel analysis such that each facility had its own intercept in the model. This was used to account for the repeated measures for each facility. The behavior of a facility level outcome was examined as a function of facility level predictors. The logit model specified a linear function at the logit (log odds) scale. The generalized linear mixed model (GLMM) can address the two challenges identified above by including facility level covariates and including longitudinal data of the facility as random effect. First, we identified the facility level attributes available in Health Facts: census region, teaching status, urban or rural status, acute status and bed size range. The second step was to include facility as random effect in the model. The third step was to calculate the Best Linear Unbiased Predictors (BLUP), estimated from the realized values of the random variables that are linear functions of the data. The computed BLUP's are unbiased as the average value of the estimate is equal to the average value of the outcome being estimated and they have minimum mean squared error within the class of linear unbiased estimators [37]. The number of sickle cell patients at each facility were included in the model to account for the different number of SCD patients observed at each facility. This was parametrized into the following model:

$$\text{Outcome} = \gamma_{00} + \gamma_{01}(\text{Number of sickle patients}) + \gamma_{02}(\text{census region}) + \gamma_{03}(\text{teaching facility}) + \gamma_{04}(\text{urban rural status}) + \gamma_{05}(\text{acute status}) + \gamma_{06}(\text{bed size range}) + \mu_{0j}$$

From this model, γ indicates the fixed effects and μ_{0j} indicates variation between facilities which represents the random effect.

Analysis

The analysis was performed using SAS statistical software version 9.4 (SAS Institute, Cary, NC, USA). The model was fit into SAS using the PROC GLIMMIX procedure with the link function equal to logit for proportion outcome [38]. The fixed effects of the number of sickle cell patients, census region, teaching facility, urban or rural status, acute status and bed size range were included in the MODEL statement after the outcome variable. Using this model, facility-specific linear unbiased percentages were generated by year. This represents the percentage of inappropriate A1c orders in each facility every year between 2011 to 2016 after accounting for the random effects inherent within the facilities. The estimated percentages for every hospital were ranked to identify the distribution of the facilities in a quartile. Two facilities were randomly chosen to observe the difference in their ranking before and after applying the GLMM.

Results

Baseline characteristics of data

Among the 863 facilities in the version of Health Facts used for this study, 393 facilities had a sickle cell population. These facilities were distributed across all 4 major geographic regions in the United States: Midwest (79 facilities), Northeast (68), South (162), and West (84). Most of the facilities (239) have a bed size less than 100, while some (133) have bed size between 100 to 500 and a few (21) have more than 500 beds. Most of the facilities (300) are urban while the rest (93) are classified as rural. The study cohort includes 37,151 sickle cell patients. The two randomly chosen facilities were large acute care facilities with a bed size greater than 500. Facility A is an urban facility in the South census region while Facility B is a rural facility in the Northeast census region. There were 1,090 SCD patients treated at facility A and 1,267 patients at facility B. The baseline characteristics of the patient population in the HF cohort and the two facilities are delineated in Table 1. Ten percent of the total patients in the HF SCD cohort (3,931

patient) had at least one A1c encounter. Facility A has a lower proportion of SCD patients with A1c encounters (15%, 168 patients) when compared to facility B (32%, 399 patients).

Table 1. Baseline characteristics of the SCD patient population in the HF cohort and across the six facilities in KCMO

Characteristics	**No. of SCD patients (%)**		
	HF	**Facility A**	**Facility B**
Patients without A1c	33,224(89)	922 (85)	868 (68)
Patients with A1c	3,927(11)	168 (15)	399 (32)
Median Age (IQR)	26 (12-47)	25 (20-38)	33 (24-45)
Gender			
Male	16,235 (44)	505 (46)	487 (38)
Female	20,821 (56)	585 (54)	780 (62)
Race			
African American	23,810 (64)	990 (91)	965 (76)
Caucasian	8,951 (24)	59 (5)	124 (10)
Other	4,390 (12)	41 (4)	178 (14)
Year			
2011	8,331 (22)	262 (24)	381 (30)
2012	8,643 (23)	313 (29)	501 (40)
2013	11,934 (32)	473 (43)	537 (42)
2014	13,294 (36)	478 (44)	562 (44)
2015	11,518 (31)	483 (44)	548 (43)
2016	11,092 (30)	375 (34)	529 (42)

Ranking the two random facilities among other facilities in the HF cohort

Ranking the two facilities based on the ratio of SCD patients with A1c orders shows that facility A is in the first (lower) quartile (< 25th percentile) in the years 2011 to 2013 and in the second quartile (< 50th percentile) in the years 2014 to 2016 while facility B is in the third quartile (< 75th percentile) in all the years between 2011 to 2015 and in the second quartile in the year 2016 (Table 2). The position of the facility A and facility B among the range of the ratio of SCD patients with A1c tests from all facilities in HF for the year 2016 is depicted in Figure 1 where both the facilities are in the second quartile.

Table 2: The percentile values for the quartiles based on the range of the ratio of SCD patients with A1c test in all the facilities and the ratio of SCD patients with A1c test in facility A and facility B for the years 2011 to 2016

Year	Ratio of SCD patients with A1c test				
	HF cohort			Facility A	Facility B
	25th Percentile	50th Percentile	75th Percentile		
2011	0.057	0.095	0.197	0.026	0.162
2012	0.11	0.171	0.333	0.030	0.197
2013	0.115	0.205	0.377	0.098	0.240
2014	0.133	0.25	0.5	0.144	0.274
2015	0.116	0.2	0.394	0.139	0.226
2016	0.108	0.225	0.5	0.129	0.193

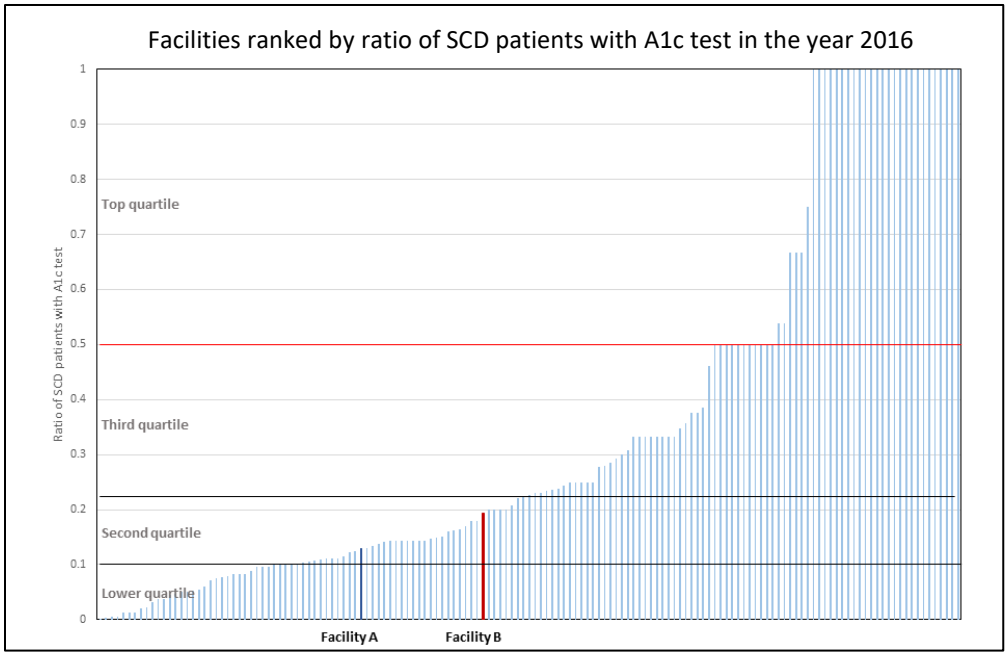

Figure 1. Facility A and facility B are within the second quartile among the range of the ratio of SCD patients with A1c tests from all facilities in HF for the year 2016

The estimated variance of the facility intercept from the empty GLMM model (without covariates) is 2.3501 with a SE of 0.1014. The fixed effect of the intercept has an estimate -2.3501 which represents the grand mean across encounters is statistically significant at p<.0001. After adding the facility level covariates, the estimated variance of the facility intercept is 1.8590 with a SE of 0.2153. By subtracting the total variance between the null model and predictor model it was determined that the predictors explain 21% of the total variance. After adjustment of confounders, facilities in the south census region are 75% less likely to have inappropriate A1c order when compared to the west region (OR, 0.251; 95% CI, 0.150-0.418; p<.00001). Acute care facilities are 72% less likely to have inappropriate A1c orders than Non-Acute care facilities (OR, 0.280; 95% CI, 0.127-0.618; p = 0.0017). Small facilities (bed size <5) are 83% less likely to have inappropriate A1c orders than large facilities (bed size > 500) (OR, 0.173; 95% CI, 0.062-0.486; p = 0.0009) (Table 3).

Table 3. Fixed effects in Generalized Linear Mixed Model (GLMM) for the relationship between facility level predictors and inappropriate A1c test orders in SCD patients

Facility characteristics		**AOR (95% CI)**	**P-value**
Number of SCD patients		1.000 (0.999-1.000)	0.1109
Census Region (Ref- Northeast)			
	Midwest	0.791 (0.450-1.390)	0.4142
	South	0.251 (0.150-0.418)	<.0001
	West	1.108 (0.617-1.988)	0.7318
Teaching status (Ref – Non-Teaching)		0.783 (0.475-1.291)	0.3376
Rural status	(Ref- Urban)	0.726 (0.468-1.126)	0.1525
Acute status	(Ref- Non-Acute)	0.280 (0.127-0.618)	0.0017
Bed Size	(Ref – 500+)		
	<5	0.173 (0.062-0.486)	0.0009
	05-99	1.150 (0.495-2.674)	0.7446
	100-199	1.102 (0.491-2.473)	0.8139
	200-299	1.022 (0.470-2.221)	0.9558
	300-499	0.943 (0.415-2.145)	0.8894

Ranking the BLUP generated for each facility by using the solution for the random effects to estimate the prediction probability for inappropriate A1c orders shows that facility A is less than the 50th percentile value and facility B is greater than the 75th percentile value for all the years between 2011 to 2016 (Table 4). The position of the facility A and facility B among the range of BLUP values estimated through GLMM from all the facilities in HF for the year 2016 is depicted in Figure 2 where facility A remains unchanged and present in the second quartile, which signifies that over-utilization of inappropriate A1c orders is a lower priority in this facility. Whereas, facility B is in the last quartile which signifies that this quality concern is a higher priority in this facility.

Table 4: The percentile values for the quartiles based on the range of the BLUP values estimated from GLMM in all the facilities and BLUP value of facility A and facility B for the years 2011 to 2016

Year	Ratio of SCD patients with A1c test				
	HF cohort			Facility A	Facility B
	25th Percentile	50th Percentile	75th Percentile		
2011	0.041	0.098	0.182	0.092	0.217
2012	0.035	0.093	0.174	0.092	0.211
2013	0.038	0.096	0.018	0.090	0.209
2014	0.037	0.104	0.208	0.090	0.208
2015	0.034	0.095	0.187	0.090	0.209
2016	0.037	0.097	0.185	0.092	0.210

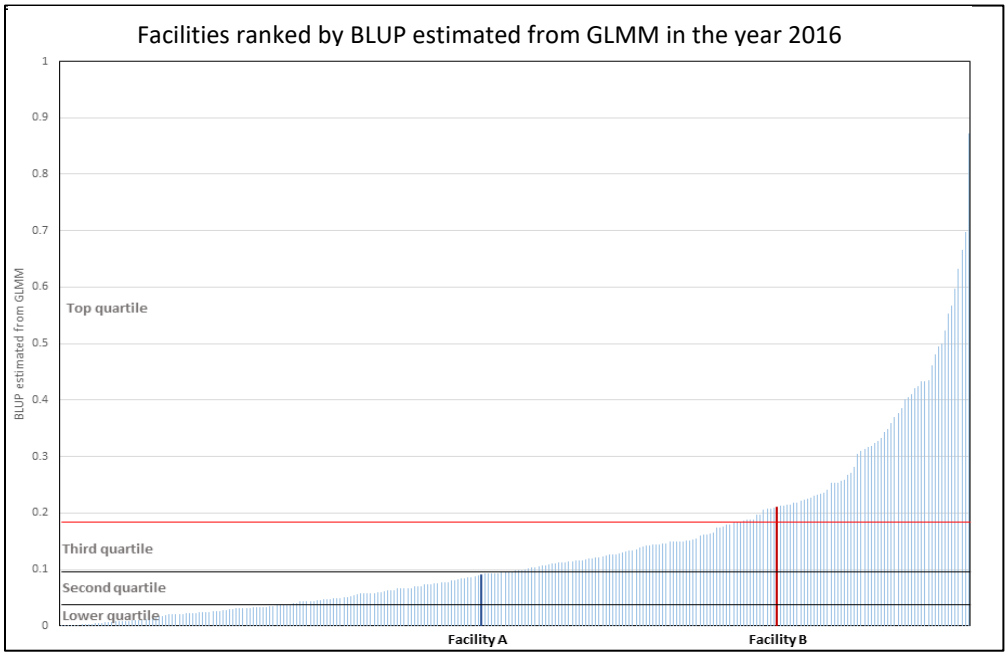

Figure 2. Facility A is in the second quartile and facility B is in the top quartile among the range of BLUP values estimated through GLMM from all the facilities in HF for the year 2016

Discussion

Using comparative multi-site EHR data to prioritize QI projects offers a powerful strategy to help members of a practice understand where their performance falls in comparison to others for topics that are not represented in widely available national data sets. As a proof of concept, we evaluated A1c test orders for sickle cell disease patients in two randomly selected facilities.

There are different techniques which local health provider organizations use to identify and prioritize QI concerns. CDC has provided a list of formalized techniques that can facilitate an orderly process [39]. Other standardized methods such as priority matrix, Hanlon method and analytic hierarchy process techniques use priority scores to

considerably narrow down the priority of the QI concerns [40,41]. However, measuring the A1c tests in SCD patients as a quality concern in the national context using a multi-site EHR provides a complementary external dimension that generates an outward perspective which surpasses preliminary prioritization techniques [42]. It not only identifies quality gaps, but also provides a deep understanding of the level of skill and processes that are required comparable to that of superior performance.

In order to apply this strategy, it was necessary to account for variations between the sites contributing the data set. We noted that many sites in our cohorts had low numbers of eligible patients or low patient volume compared to other facilities. GLMM and BLUP have been demonstrated to be useful for similar challenges in ranking and selection in the context of animal [43], plant breeding [44], to predict an individual's risk of developing cancer [45] and in genetics [37] after taking into account of variation associated with the environmental factors. The quality measurement plan (QMP), which regards probability of distribution of an outcome as the true quality index, is an example of BLUP where non-normal distributions are assumed [46]. It is a useful technique when the ideal ranking involves random effects. Ranking the predicted probabilities of BLUP estimated from GLMM for all facilities offers a novel solution to this issue as it accounts for the variability associated with each level of the outcome by including facility attributes as covariates.

Bias due to the variations between the facilities in the EHR was observed when the facilities were ranked based only on the ratio of SCD patients with A1c orders over the total number of SCD patients within that facility per year. There were many facilities with a higher or lower ratio irrespective of the number of patients in the facility. For example, in the year 2016, 122 facilities had a ratio of 0 which indicates A1c tests are not ordered for SCD patients in those facilities. However, the number of SCD patients present in those 122 facilities ranged between 1 to 194 which should have been considered and appropriately ranked. A facility with 194 SCD patients and without a single inappropriate A1c order should be ranked higher than a facility with just 1 SCD patients and no A1c test order. There were also 26 facilities with a ratio of 1, for which the range of SCD patients was between 1 to 11. Furthermore, there may be considerable variation in the observations due to the differences in the characteristics between facilities such as variation in the type of specialty status, geographic location and size of the facility, which were not accounted for when the unadjusted ratio data was used to rank the facilities. Due to these limitations, ranking the facilities based on the ratio of the outcome could be distorted where overutilization of A1c orders at Facility B may not be considered a high priority quality gap as it is in the third quartile.

The extent to which facility characteristics predict inappropriate A1c orders was determined by the application of GLMM. Facility level predictors correlate with 21% of the total variance. Census region, bed size range and acute status were significant risk factors. The significance of the BLUP over the ratio of outcome can be observed when the GLMM assigns different BLUP values to facilities with the same ratio based on their facility characteristics and number of patients/encounters. Facility A and B with a similar bed size range, were in the same second quartile for the year 2016 when the facilities were ranked based on the ratio of SCD patients with A1c test. This is because of the high number of facilities with a ratio of zero and one which were properly segregated based on the number of SCD patients and facility characteristics. For example, in the year 2016, a facility with only 1 SCD patient without any A1c test has the same ratio of 0, as does a facility with 194 SCD patients without any A1c tests. However, the BLUP value for the latter facility is 0.00013 and is ranked higher than the prior facility with a BLUP value of 0.113. From another example, among two facilities with a similar SCD population (180 SCD patients) and no A1c encounters, the larger facility (bed size > 100) is ranked higher than the smaller facility (bed size < 100) in all six years which demonstrates that this model accounts for facility level characteristics to assign ranks.

Ranking the predicted probabilities estimated from BLUP of all the facilities shows that the QI initiative for A1c test orders for sickle cell patients in facility B should have a higher priority (fourth quartile) than compared to facility A (second quartile), if national context was the primary factor in the prioritization (Figure 2). This strategy addresses the limitations associated with the previous model by adjusting the outcome measure between and within facilities. The outcome measures are meaningful only when adequately adjusted for facility level characteristics and the differences between the patient and procedural characteristics of the facilities. In this application, the data structures are longitudinal (i.e., across multiple facility encounters) and the analyses involve repeated measures at several years which are modelled by GLMM. It accounts for the non-normal distribution of the dependent variable, its restricted range, the relation between mean and variance and includes both fixed and random effects of the varying covariates [47]. The BLUP methodology uses the solution for the random effects in GLMM and produces the best estimator of probabilities to predict the outcome variable in a facility [48].

Based on the GLMM analysis, Facility B would prioritize reduction of A1c tests in SCD patients while facility A may not. In addition to facility B, there are 68 facilities in the HF cohort for the year 2016 which are ranked in the bottom quartile with respect to inappropriate A1c orders, suggesting that these two quality concerns are prevalent in many facilities. Further review of this data could provide insights into the institutional and provider behaviors associated with these ordering patterns. Additional confounders may include differences in experience and practice in providing A1c testing among Health Facts facilities, or perhaps a higher proportion of patients at risk for diabetes are acquired at these bottom quartile facilities than other healthcare entities. However, provider characteristics and comorbidities of the SCD patients were not included in this model. Another limitation is analyzing de-identified EHR data is challenging as there might be errors in capturing the complete SCD cohort. There is no way to verify the accuracy of diagnostic codes (ICD-9-CM, ICD-10-CM) used in Health Facts which are for administrative purposes and hence, are unreliable when compared to gold standard clinician extraction [49,50]. Our findings suggest the need for interventions to reduce or eliminate the use of A1c tests in SCD patients for the diagnoses or management of diabetes.

Conclusion

Analyzing multi-institutional EHR data, after adjusting the bias due to the covariates by using GLMM to generate BLUP's, provided a useful way to evaluate the severity of QI project (Figure 2). The GLMM in this study offered several advantages. First, the linear unbiased predicted percentages from GLMM adjusted the bias due to covariates. Second, GLMMs were able to model the longitudinal structure of the data. Third, by including the facility as a random variable in the mixed model, we were able to generalize the inference of the fixed effect to the population. Additionally, one favorable property of the BLUP generated from the GLMM is shrinkage towards the mean, which anticipates regression of the facility characteristics to the mean of the outcome observed by every facility. Finally, we developed an analytic pipeline which can be easily adopted to different measurements of outcome variables, such as ordinal or nominal outcome. While our focus was prioritizing QI projects between facilities, the methods described here are useful for any comparison of diverse facilities in a multi-contributor EHR data warehouse.

Acknowledgments

This work was funded by CDC Cooperative Agreement NU47OE000105-01-01. We wish to acknowledge the contributions of Suman Sahil from University of Missouri-Kansas City as well as Dr. Kamani Lankachandra from Truman Medical Center for early involvement in this project

References

1. Lohr KN. Medicare: a strategy for quality assurance. Vol. 1. National Academies Press; 1990.
2. Plebani M, Sciacovelli L, Marinova M, Marcuccitti J, Chiozza ML. Quality indicators in laboratory medicine: a fundamental tool for quality and patient safety. Clin Biochem. 2013;46(13–14):1170–4.
3. Medicare C for, Baltimore MS 7500 SB, Usa M. Overview [Internet]. 2018 [cited 2019 Jan 30]. Available from: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityInitiativesGenInfo/index.html
4. NQF: Improving Healthcare Quality [Internet]. [cited 2019 Jan 30]. Available from: https://www.qualityforum.org/Setting_Priorities/Improving_Healthcare_Quality.aspx
5. Muething SE, Goudie A, Schoettker PJ, Donnelly LF, Goodfriend MA, Bracke TM, et al. Quality Improvement Initiative to Reduce Serious Safety Events and Improve Patient Safety Culture. Pediatrics. 2012 Aug;130(2):e423–31.
6. Institute for Healthcare Improvement: The IHI Triple Aim [Internet]. [cited 2019 Jan 30]. Available from: http://www.ihi.org:80/Engage/Initiatives/TripleAim/Pages/default.aspx
7. NQF: Priority Setting for Healthcare Performance Measurement: Addressing Performance Measure Gaps for Adult Immunizations [Internet]. [cited 2019 Jan 30]. Available from: http://www.qualityforum.org/Publications/2014/08/Priority_Setting_for_Healthcare_Performance_Measurement__Addressing_Performance_Measure_Gaps_for_Adult_Immunizations.aspx
8. Medicare C for, Baltimore MS 7500 SB, Usa M. Core Measures [Internet]. 2017 [cited 2019 Jan 30]. Available from: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityMeasures/Core-Measures.html
9. Safety I of M (US) C on DS for P, Aspden P, Corrigan JM, Wolcott J, Erickson SM. Adverse Event Analysis [Internet]. National Academies Press (US); 2004 [cited 2018 Oct 20]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK216102/

10. Heavner JJ, Siner JM. Adverse Event Reporting and Quality Improvement in the Intensive Care Unit. Clin Chest Med. 2015 Sep;36(3):461–7.
11. Rafter N, Hickey A, Condell S, Conroy R, O'Connor P, Vaughan D, et al. Adverse events in healthcare: learning from mistakes. QJM Int J Med. 2015 Apr 1;108(4):273–7.
12. Rumball-Smith J, Shekelle PG, Bates DW. Using the Electronic Health Record to Understand and Minimize Overuse. JAMA. 2017 17;317(3):257–8.
13. Waitman LR, Warren JJ, Manos EL, Connolly DW. Expressing Observations from Electronic Medical Record Flowsheets in an i2b2 based Clinical Data Repository to Support Research and Quality Improvement. AMIA Annu Symp Proc. 2011;2011:1454–63.
14. Himes BE, Dai Y, Kohane IS, Weiss ST, Ramoni MF. Prediction of Chronic Obstructive Pulmonary Disease (COPD) in Asthma Patients Using Electronic Medical Records. J Am Med Inform Assoc. 2009 May 1;16(3):371–9.
15. Sittig DF, Hazlehurst BL, Brown J, Murphy S, Rosenman M, Tarczy-Hornoch P, et al. A survey of informatics platforms that enable distributed comparative effectiveness research using multi-institutional heterogeneous clinical data. Med Care. 2012 Jul;50(Suppl):S49–59.
16. Evans RS, Lloyd JF, Pierce LA. Clinical Use of an Enterprise Data Warehouse. AMIA Annu Symp Proc. 2012 Nov 3;2012:189–98.
17. NIS Database Documentation [Internet]. [cited 2019 Jan 31]. Available from: https://www.hcup-us.ahrq.gov/db/nation/nis/nisdbdocumentation.jsp
18. KID Database Documentation [Internet]. [cited 2019 Jan 31]. Available from: https://www.hcup-us.ahrq.gov/db/nation/kid/kiddbdocumentation.jsp
19. NEDS Overview [Internet]. [cited 2019 Jan 31]. Available from: https://www.hcup-us.ahrq.gov/nedsoverview.jsp
20. PHIS [Internet]. [cited 2019 Jan 31]. Available from: https://www.childrenshospitals.org/phis
21. ACS National Surgical Quality Improvement Program [Internet]. American College of Surgeons. [cited 2019 Jan 31]. Available from: https://www.facs.org/quality-programs/acs-nsqip
22. DeShazo JP, Hoffman MA. A comparison of a multistate inpatient EHR database to the HCUP Nationwide Inpatient Sample. BMC Health Serv Res. 2015 Sep 15;15:384.
23. Strack B, DeShazo JP, Gennings C, Olmo JL, Ventura S, Cios KJ, et al. Impact of HbA1c measurement on hospital readmission rates: analysis of 70,000 clinical database patient records. BioMed Res Int. 2014;2014:781670.
24. Petrick JL, Nguyen T, Cook MB. Temporal trends of esophageal disorders by age in the Cerner Health Facts database. Ann Epidemiol. 2016 Feb;26(2):151-154.e4.
25. Yang H, Chaudhari P, Zhou Z-Y, Wu EQ, Patel C, Horn DL. Budget impact analysis of liposomal amphotericin B and amphotericin B lipid complex in the treatment of invasive fungal infections in the United States. Appl Health Econ Health Policy. 2014 Feb;12(1):85–93.
26. Shorr AF, Myers DE, Huang DB, Nathanson BH, Emons MF, Kollef MH. A risk score for identifying methicillin-resistant Staphylococcus aureus in patients presenting to the hospital with pneumonia. BMC Infect Dis. 2013 Jun 6;13:268.
27. Smaldone A. Glycemic control and hemoglobinopathy: when A1C may not be reliable. Diabetes Spectr. 2008;21(1):46–9.
28. Lippi G, Targher G. Glycated hemoglobin (HbA1c): old dogmas, a new perspective? Clin Chem Lab Med. 2010 May;48(5):609–14.
29. Thom CS, Dickson CF, Gell DA, Weiss MJ. Hemoglobin Variants: Biochemical Properties and Clinical Correlates. Cold Spring Harb Perspect Med [Internet]. 2013 Mar [cited 2018 Oct 11];3(3). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3579210/
30. Bry L, Chen PC, Sacks DB. Effects of hemoglobin variants and chemically modified derivatives on assays for glycohemoglobin. Clin Chem. 2001 Feb;47(2):153–63.
31. Speeckaert M, Van Biesen W, Delanghe J, Slingerland R, Wiecek A, Heaf J, et al. Are there better alternatives than haemoglobin A1c to estimate glycaemic control in the chronic kidney disease population? Nephrol Dial Transplant Off Publ Eur Dial Transpl Assoc - Eur Ren Assoc. 2014 Dec;29(12):2167–77.
32. Association AD. Standards of medical care in diabetes—2014. Diabetes Care. 2014;37(Supplement 1):S14–80.
33. Sickle Cell Trait & Other Hemoglobinopathies & Diabetes (For Providers) | NIDDK [Internet]. National Institute of Diabetes and Digestive and Kidney Diseases. [cited 2018 Feb 16]. Available from: https://www.niddk.nih.gov/health-information/diagnostic-tests/sickle-cell-trait-hemoglobinopathies-diabetes

34. NGSP: Factors that Interfere with HbA1c Test Results [Internet]. [cited 2017 May 24]. Available from: http://www.ngsp.org/factors.asp
35. Michalik DE, Panepinto JA| Medical College of Wisconsin| Sickle Cell Disease| PheKB [Internet]. 2017 [cited 2020 Dec 15]. Available from: https://phekb.org/phenotype/615
36. Robinson GK. That BLUP is a Good Thing: The Estimation of Random Effects. Stat Sci. 1991;6(1):15–32.
37. Schabenberger O. Introducing the GLIMMIX procedure for generalized linear mixed models. SUGI 30 Proc. 2005;196.
38. CDC - National Public Health Performance Standards - STLT Gateway [Internet]. 2018 [cited 2019 Feb 15]. Available from: https://www.cdc.gov/od/ocphp/nphpsp/documents/Prioritization.pdf
39. Wang H, Xie M, Goh TN. A comparative study of the prioritization matrix method and the analytic hierarchy process technique in quality function deployment. Total Qual Manag. 1998;9(6):421–30.
40. Rowan L, Walsh R, Durrant L. Creating national emergency pediatric health care priorities through a modified nominal group process and Hanlon method. In: Proceedings of the 132nd Annual American Public Health Association Meetings. 2004.
41. Hutton R. Effective benchmarking through a prioritization methodology. Total Qual Manag. 1995;6(4):399–412.
42. Tavernier A. Advantages of BLUP animal model for breeding value estimation in horses. Livest Prod Sci. 1988;20(2):149–60.
43. Piepho HP, Möhring J, Melchinger AE, Büchse A. BLUP for phenotypic selection in plant breeding and variety testing. Euphytica. 2008 May 1;161(1):209–28.
44. Martínez-Ávila JC, Guillén-Ponce C, Earl J, García-Cortés LA. Hereditary Lifetime Cancer Risk Assessment Modeling: A Case Study in Breast Cancer. Int J Mol Genet Gene Ther. 2016;1(2).
45. Hoadley B. The quality measurement plan (QMP). Bell Syst Tech J. 1981;60(2):215–73.
46. McCulloch CE, Neuhaus JM. Generalized linear mixed models. Wiley StatsRef Stat Ref Online. 2014;
47. Henderson CR. Best linear unbiased estimation and prediction under a selection model. Biometrics. 1975;423–47.
48. Quan H, Li B, Saunders LD, Parsons GA, Nilsson CI, Alibhai A, et al. Assessing Validity of ICD-9-CM and ICD-10 Administrative Data in Recording Clinical Conditions in a Unique Dually Coded Database. Health Serv Res. 43(4):1424–41.
49. O'Malley KJ, Cook KF, Price MD, Wildes KR, Hurdle JF, Ashton CM. Measuring Diagnoses: ICD Code Accuracy. Health Serv Res. 40(5p2):1620–39.