*MEDINFO 2021: One World, One Health – Global Partnership for Digital Innovation*
*P. Otero et al. (Eds.)*


*doi:10.3233/SHTI220104*

# Testing a Novel Inpatient Respiratory Depression Electronic Clinical Quality Measure (eCQM) for Orthopedic Practice in Two Large U.S. Health Systems

**Ania Syrowatka[a,b], Troy Li[a], Mica Curtin-Bowen[a], Avery Pullman[a], Stuart R. Lipsitz[a,b], Michael Sainlaire[a], Wenyu Song[a,b], Tien Thai[a], Alexandra Businger[a], Kevin J. Bozic[c], William A. Jiranek[d], Jay R. Lieberman[e], David W. Bates[a,b], Patricia C. Dykes[a,b]**

*[a]Department of General Internal Medicine, Brigham and Women's Hospital, Boston, MA, USA*
*[b]Harvard Medical School, Harvard University, Boston, MA, USA*
*[c]Dell Medical School, The University of Texas at Austin, Austin, TX, USA*
*[d]Department of Orthopaedic Surgery, Duke University, Durham, NC, USA*
*[e]Keck School of Medicine, University of Southern California, Los Angeles, CA, USA*

**Abstract**

*The objective of this study was to assess the feasibility of using an electronic clinical quality measure (eCQM) to assess inpatient respiratory depression rates following elective primary total hip or total knee arthroplasty using data routinely collected in electronic health records. Measure testing was conducted at two large urban, academic health systems – Mass General Brigham and a geographically distant system in southern U.S. The risk-adjusted inpatient respiratory depression rates were 3.83 and 2.73% for the two health systems, respectively. Clinician group rates ranged from 1.40 to 4.35%, demonstrating opportunity for improvement. Both the data and measure specifications showed strong reliability and validity to allow for calculation of accurate and comparable rates of inpatient respiratory depression.*

***Keywords:***

Patient safety, orthopedics, electronic health records.

## Introduction

Postoperative respiratory depression is a serious event that places patients at risk of hypoxia, anoxia, severe brain damage, cardiac arrest, and death [1]. It has been associated with longer hospital stays (by 55%), higher healthcare costs (by 47%), increased 30-day readmission rates (by 36%), and a 3.4 times higher rate of mortality in patients receiving opioids [2]. Opioid-induced respiratory depression is an important contributor to postoperative respiratory depression [3]. Other contributors to respiratory depression include residual anesthesia, muscle paralysis, use of sedatives, and comorbid lung diseases.

Key challenges to accurate and comparable measurement of respiratory depression are lack of a universal definition, variability in measurement approaches, and heterogeneous populations. There are no reliable diagnostic codes to document the occurrence of respiratory depression; rather codes focus on related outcomes (e.g., hypoxemia and respiratory arrest). Furthermore, respiratory depression can be evaluated using a variety of methods including: respiratory rate, oxygen saturation (SP02), partial pressure of carbon dioxide in arterial blood, exhaled end-tidal carbon dioxide, and heart rate below or above certain thresholds [3]. However, even when using the same method, different thresholds may be applied to define the presence of respiratory depression (e.g., SP02 $\leq$ 90% vs. $\leq$ 88%), making estimates of the incidence of respiratory depression less comparable between groups. Metrics using data routinely documented in electronic health records (EHRs) could help to overcome some of these challenges and allow for more accurate measurement by leveraging data elements not available in claims data such as vital signs.

Our team at Brigham and Women's Hospital was under agreement with the Centers for Medicare and Medicaid Services (CMS) to develop and test novel electronic clinical quality measures (eCQMs). The objective of this work was to develop and test an eCQM that reports on the inpatient respiratory depression rate following total hip arthroplasty (THA) or total knee arthroplasty (TKA). The development process addressed the variability described above to create a standardized measurement approach. This paper describes the testing of the eCQM in two large U.S. health systems that use different EHR vendors.

## Methods

### Study cohorts

Testing of the eCQM was conducted using data routinely collected in EHRs. The target populations of adult patients aged 18+ years who received elective primary THA or TKA were extracted from the Mass General Brigham (MGB) Epic EHR from January 1, 2016 to December 31, 2019, and a southern U.S. system Cerner EHR from January 1, 2017 to December 31, 2019. Descriptive statistics were calculated to characterize the sociodemographic information of the cohorts. The target population was used to assess data reliability and validity.

Data documented in the EHRs (e.g., inpatient vs. outpatient procedures) were used to define the denominator population for calculation of eCQM rates as well as assessment of the reliability and validity of the measure specifications.

#### *Rates of inpatient respiratory depression*

Measure development was informed by the published literature, review of existing clinical quality measures related to respiratory depression as well as those under development, consultations with a Technical Expert Panel, and interviews with patients, orthopedic providers and payers. The following specifications were developed:

- Denominator: adult patients aged 18+ years who received an elective primary THA or TKA covered by any healthcare payer. The measure exclusion criteria were harmonized with existing orthopedic clinical quality measures [4] and also excluded patients:
  - Discharged against medical advice
  - With more than two THA or TKA procedure codes documented during the hospital stay
  - Who received an outpatient procedure
- Numerator: the subset of patients from the denominator who experienced postoperative respiratory depression during the inpatient stay as defined by at least one of the following documented in the EHR:
  - Diagnostic code for respiratory depression-related outcome or respiratory failure
  - Procedure code for mechanical ventilation or intubation after the procedure
  - At least two SP02 readings ≤ 88% and > 30% within a period of 24 hours
  - At least three SP02 readings ≤ 88% and > 30% during the inpatient stay

Unadjusted and risk-adjusted eCQM inpatient respiratory depression rates were calculated for both health systems as well as for each clinician group. Rates were risk adjusted based on patient age, sex, race, primary language, zip code (proxy for household income), type of insurance, orthopedic procedure (i.e., THA or TKA), body mass index, smoking status, and relevant comorbidities. eCQM rates are reported as percentages.

The risk adjustment model used to calculate the predicted over expected (P/E) ratios is presented in Appendix Table 1. The P/E ratios were multiplied by the unadjusted rate of the total sample to estimate the risk-adjusted rates.

The interclass correlation coefficient (ICC) was calculated to assess the variation in inpatient respiratory depression rates that was explained by clinician groups.

***Data reliability and validity***

Data reliability and validity were evaluated to determine the accuracy and feasibility of calculating the eCQM using routinely collected EHR data from the two largest U.S. vendors, namely Epic and Cerner.

The availability of data elements required for measure calculation and risk adjustment was assessed, and reliability was evaluated using the National Quality Forum Feasibility Scorecard [5].

Data validity was assessed through manual chart review of a random sample of patients from MGB. Percentage agreement and Kappa scores were calculated to quantify agreement with the eCQM, based on classification of patients as included in the 'numerator' or 'denominator only,' or 'excluded' from measure calculation.

***Measure reliability and validity***

Data were subsequently randomly split 50:50 into test and validation samples. The risk-adjustment model presented in Appendix Table 1 was developed using hierarchical logistic regression based on the test data and applied to the validation sample to assess the reliability.

P/E ratios were calculated for the test and validation samples and multiplied by the unadjusted rate of the total sample to estimate the risk-adjusted rates. The P/E ratios were ranked in the test and validation samples and agreement was assessed using Spearman's rank-order correlation coefficient [6]. The C-statistic and Hosmer-Lemeshow test were used to assess the goodness-of-fit of the logistic model [7].

## Results

**Study cohorts**

The demographic characteristics of all patients who received an elective primary THA or TKA at the MGB or southern U.S. system are provided in Table 1. A subset of 16,428 patients from MGB and 11,304 from the southern U.S. system met the denominator criteria and were included for measure calculation.

*Table 1– Characteristics of the Target Population*

| Demographic information | MGB system n=17,324 | South U.S. system n=11,881 |
|---|---|---|
| Age, mean | 66.05 | 65.63 |
| Age ≥ 65 years, % | 57.57 | 56.93 |
| 18 ≤ age ≤ 65 years, % | 42.43 | 43.07 |
| Male, % | 42.12 | 42.03 |
| Body mass index, median | 29.61 | 30.47 |
| White, % | 89.87 | 68.02 |
| Black/African American, % | 3.65 | 10.98 |
| Hispanic, % | 0.56 | N/A |
| English as first language, % | 95.41 | 92.32 |
| Smoker, % | 5.56 | 6.25 |
| Public insurance, % | 46.78 | 56.62 |
| Income, median, USD | 72,935 | 63,795 |

*N/A*, not applicable – data not collected in EHR.

The two cohorts were similar with regard to age and sex. The southern U.S. cohort had greater racial diversity, a higher percentage of patients covered by public insurance, and a lower median income.

Both health systems showed a similar breakdown by orthopedic procedures. At MGB, 45.42% of patients received a THA and 54.58% underwent TKA. At the southern U.S. system, 48.85% and 51.15% of patients received a THA and TKA, respectively.

***Rates of inpatient respiratory depression***

There were six orthopedic clinician groups at MGB, and the overall risk-adjusted inpatient respiratory depression rate was 3.83% with clinician-group rates ranging from 2.84 to 4.35% (Table 2). The southern U.S. system had 11 clinician groups with an overall risk-adjusted rate of 2.73% and clinician-group rates between 1.40 and 4.30%.

The P/E ratios ranged from 0.44 to 1.36 (a three-fold difference) across all clinician groups. The ICC was 0.0692, indicating that 6.92% of the variation in eCQM rates was explained by clinician groups.

*Table 2– Unadjusted and Risk-adjusted eCQM Rates for Model Development (Test) Sample, Overall and by Clinician Group for each Health System*

| Health system, clinician group | Unadjusted rate, % | P/E ratio | Adjusted rate, % |
|---|---|---|---|
| MGB, overall | 3.30 | 1.20 | 3.83 |
| MGB, A | 3.31 | 1.36 | 4.35 |
| MGB, B | 5.13 | 1.34 | 4.29 |
| MGB, C | 3.40 | 1.32 | 4.21 |
| MGB, D | 6.67 | 1.20 | 3.82 |
| MGB, E | 3.82 | 0.92 | 2.95 |
| MGB, F | 2.47 | 0.89 | 2.84 |
| South U.S., overall | 3.05 | 0.85 | 2.73 |
| South U.S., A | 1.32 | 0.91 | 2.90 |
| South U.S., B | 2.85 | 0.97 | 3.11 |
| South U.S., C | 11.76 | 0.97 | 3.10 |
| South U.S., D | 0.84 | 0.44 | 1.40 |
| South U.S., E | 3.89 | 0.90 | 2.87 |
| South U.S., F | 3.31 | 0.92 | 2.94 |
| South U.S., G | 11.46 | 1.35 | 4.30 |
| South U.S., H | 6.29 | 1.32 | 4.21 |
| South U.S., I | 3.32 | 0.84 | 2.70 |
| South U.S., J | 2.20 | 0.79 | 2.51 |
| South U.S., K | 3.21 | 1.02 | 3.27 |

***Data reliability and validity***

Data required for measure calculation and risk adjustment were routinely documented in EHRs. Manual chart review of 230 randomly selected patients at MGB confirmed that the data required for measure calculation were available in the EHR. Overall, the MGB EHR captured 99.28% of the necessary data for risk adjustment, and the southern U.S. EHR contained 98.33%. However, data on smoking status was missing for 12.94% of patients who received their procedure within the southern U.S. system (Table 3). Assessments based on the Feasibility Scorecard showed that all required data elements were available, accurate, coded using nationally accepted terminology standards, and routinely documented during care at both health systems.

*Table 3– Data Element Availability*

| Variable | Missing data elements for MGB system, % | Missing data elements for southern U.S. system, % |
|---|---|---|
| Insurance type | 0.10 | 0.00 |
| Body mass index | 0.72 | 0.02 |
| Primary language | 1.12 | 0.15 |
| Smoking status | 1.41 | 12.94 |
| Zip code | 0.14 | 0.15 |
| Sex | 0.00 | 0.06 |
| Race | 2.33 | 0.04 |
| Admit age | 0.00 | 0.00 |
| Condition | None of the cases had all NULL values for any of the condition columns. | |

In classifying patients as included in the 'numerator' or 'denominator only,' or 'excluded' from measure calculation, the percentage agreement was 97.50% in round 5 of chart review with a corresponding Kappa of 0.96 (Table 4), based on a manual chart review of samples of patients from the MGB system. This indicated excellent agreement between manual and eCQM results.

Given the high percentage agreement and Kappa in round 1 of the chart review, no corrections to the measure specifications were necessary. In the few cases of disagreement between chart review and the eCQM, it was determined that the eCQM classifications were correct.

*Table 4– Agreement between Manual Chart Review and eCQM Classification of Patients*

| Chart review round | n | % Agreement | Kappa |
|---|---|---|---|
| 1 | 39 | 97.44 | 0.96 |
| 2 | 41 | 100.00 | 1.00 |
| 3 | 40 | 90.00 | 0.84 |
| 4 | 40 | 92.50 | 0.88 |
| 5 | 40 | 97.50 | 0.96 |
| Exclusion only | 60 | 95.50 | 0.90 |

***Measure reliability and validity***

The measure showed strong reliability using test and validation samples (Table 5). The P/E ratios for both samples resulted in similar rankings of the 17 clinician groups with a Spearman's rank-order correlation of 0.77.

The model showed good discrimination between patients who experienced inpatient respiratory depression and patients who did not. The C-statistic was 0.78 in the test sample and 0.77 in the validation sample. A C-statistic of 0.5 indicates performance equal to random chance, and C-statistics > 0.70 indicate good model performance.

The Hosmer-Lemeshow goodness-of-fit test had p-values of 0.56 and 0.98 in the test and validation samples, respectively. P-values > 0.10 indicate good fit between the observed and expected results.

*Table 5– Comparison between Predicted/Expected (P/E) Ratios and Risk-adjusted eCQM Rates for Test and Validation Samples by Clinician Group*

| Health system, clinician group | P/E ratio (test) | Adj. rate, % (test) | P/E ratio (valid) | Adj. rate, % (valid) |
|---|---|---|---|---|
| MGB, A | 1.36 | 4.35 | 1.37 | 4.38 |
| MGB, B | 1.34 | 4.29 | 1.44 | 4.60 |
| MGB, C | 1.32 | 4.21 | 1.47 | 4.69 |
| MGB, D | 1.20 | 3.82 | 1.97 | 6.28 |
| MGB, E | 0.92 | 2.95 | 0.96 | 3.06 |
| MGB, F | 0.89 | 2.84 | 0.94 | 2.99 |
| South U.S., A | 0.91 | 2.90 | 0.80 | 2.56 |
| South U.S., B | 0.97 | 3.11 | 0.85 | 2.73 |
| South U.S., C | 0.97 | 3.10 | 1.06 | 3.40 |
| South U.S., D | 0.44 | 1.40 | 0.41 | 1.32 |
| South U.S., E | 0.90 | 2.87 | 0.74 | 2.36 |
| South U.S., F | 0.92 | 2.94 | 1.03 | 3.29 |
| South U.S., G | 1.35 | 4.30 | 1.20 | 3.84 |
| South U.S., H | 1.32 | 4.21 | 1.25 | 4.01 |
| South U.S., I | 0.84 | 2.70 | 1.07 | 3.42 |
| South U.S., J | 0.79 | 2.51 | 0.72 | 2.32 |
| South U.S., K | 1.02 | 3.27 | 0.92 | 2.95 |

*Adj*, adjusted; *Valid*, validation.

## Discussion

Development of this eCQM addressed many challenges associated with measurement of inpatient respiratory depression by leveraging data elements, specifically vital signs, routinely documented in EHRs but not available in administrative health data. This approach allows for more accurate estimates and fair comparisons of performance between health systems, hospitals, and orthopedic clinician groups. This study assessed the feasibility of implementing this eCQM at two geographically distinct health systems that use different EHR vendors. The eCQM can be relatively easily implemented within Epic and Cerner EHR systems to provide automated reporting of performance to CMS.

The data required to calculate the eCQM were available, reliable and valid. This demonstrated the feasibility of using routinely collected EHR data to accurately measure inpatient respiratory depression rates using the proposed eCQM specifications.

The measure also showed strong reliability and validity through test-retest assessments with a strong Spearman's rank-order correlation of 0.77. Both the test and validation samples showed reproducible C-statistics of ~0.77, and the logistic model was considered to be a good fit for both samples.

The system-wide eCQM inpatient respiratory depression rates were 3.83 and 2.73% for MGB and the southern U.S. system, respectively. Clinician group rates ranged from 1.40 to 4.35% across both systems, highlighting opportunity for improvement.

Only 6.92% of the variation in rates was explained by the orthopedic clinician groups. However, the eCQM was tested in two large urban, high-performing health systems that are not representative of all orthopedic practices. We would expect to see greater variability in P/E ratios and rates with the inclusion of additional health systems and orthopedic clinician groups. Furthermore, inclusion of lower performing groups could increase the percentage of the variation explained by the clinician groups.

We initially developed the eCQM for the Merit-based Incentive Payment System (MIPS), which reports at the clinician and clinician group levels. However, a limitation of reporting at this level is that multiple disciplines including orthopedics, anesthesia and nursing among others may be responsible when this complication occurs – it is clearly a multidisciplinary issue, but patient selection which is done by orthopedic providers plays a key role. Given the broad range of providers involved in the prevention and management of inpatient respiratory depression, it may be more meaningful to report this measure at the hospital or health system level. Members of the Technical Expert Panel for development of this eCQM advised that it should be reported as a facility-based measure. Notably, each orthopedic provider group in our study represented one hospital, meaning that the rates reported are also applicable at the facility level. As a next step, we aim to determine the most appropriate levels of attribution for this measure, since there are various pathways in the Quality Payment Program where the proposed measure could be implemented.

Another limitation of the eCQM is that the measure population is expected to change over time as lower-risk THAs and TKAs shift to outpatient procedures, which are excluded from the measure calculation. This will undoubtedly impact the ability to interpret trends and quality improvement over time. However, there will remain a large number of clinician groups and patient populations who would still benefit from the measure and the eCQM will continue to provide strong and comparable cross-sectional assessments of performance between health systems, hospitals, and orthopedic groups. The measure may even become more meaningful as it is applied to populations of more complex and higher-risk patients who are more likely to experience inpatient respiratory depression.

These limitations serve to highlight one of the strengths of implementing this measure as an eCQM rather than a claims-based measure – access to data in real time – which is more flexible and adaptable for managing dynamic populations. The eCQM is calculated using routinely collected data, which means that health systems, hospitals and clinician groups can leverage the information contained within their local EHR in real time to assess their performance to date, track progress, and evaluate whether quality improvement activities, such as use of multimodal analgesic regimens or appropriate monitoring for respiratory depression, are having the intended effects.

## Conclusions

Both the data and eCQM specifications demonstrated strong reliability and validity to allow for calculation of accurate rates of inpatient respiratory depression following THA and TKA using routinely collected EHR data from the two largest U.S. vendors – Epic and Cerner.

The eCQM demonstrated variability in performance between orthopedic groups, highlighting opportunities to improve the quality and safety of postoperative care. The overarching goal of our work is to obtain National Quality Forum endorsement and have this eCQM implemented as part of the CMS Quality Payment Program to provide meaningful performance measurement that can be used to drive local quality improvement.

## Acknowledgements

This project was supported by grant number 1V1CMS331637-03-00 from CMS. The content is solely the responsibility of the authors and does not necessarily represent the official views of CMS.

## References

[1] A. Miskovic, A. Lumb, Postoperative pulmonary complications, *Br J Anaesth* **118** (2017), 317-334.

[2] E.R. Kessler, M. Shah, S. Gruschkus, A. Raju, Cost and quality implications of opioid-based postsurgical pain control using administrative claims data from a large health system: opioid-related adverse events and their impact on clinical and economic outcomes, *Pharmacotherapy* **33** (2013), 383-391.

[3] S. Ayad, A.K. Khanna, S.U. Iqbal, N. Singla, Characterisation and monitoring of postoperative respiratory depression: current approaches and future considerations, *Br J Anaesth* **123** (2019), 378-391.

[4] Hospital-level risk-standardized complication rate (RSCR) following elective primary total hip arthroplasty (THA) and/or total knee arthroplasty (TKA). National Quality Forum, https://www.qualityforum.org/QPS/1550.

[5] CMS, Baltimore MD: Measures Management System, Measure Management & You (2018).

[6] C. Spearman, Correlation calculated from faulty data, *Br J Psychol* **3** (1910), 271-295.

[7] D.W. Hosmer, S.A. Lemeshow, R.X. Sturdivant, Applied Logistic Regression, *Wiley* (2013).

**Address for correspondence**

Ania Syrowatka: asyrowatka@bwh.harvard.edu
Patricia C. Dykes: pdykes@bwh.harvard.edu

## Appendix

*Appendix Table 1– Risk Adjustment Variables and Estimated Logistic Regression Coefficients*

| Effect | Coefficient |
|---|---|
| Intercept | –8.6879 |
| Age | 0.03694 |
| Age*age | –0.00008 |
| Median_income_1000 | –0.00733 |
| Median_income*median_income | 0.000056 |
| Private | 0.1911 |
| Female | –0.06430 |
| Hip | 0.1319 |
| Black | –0.2185 |
| White | 0.03143 |
| English | 0.5357 |
| Smoke | 0.3293 |
| Body mass index | 0.07404 |
| Body mass index*body mass index | –0.00014 |
| Bone/ joint/ muscle infections/ necrosis | 0.1442 |
| Cardio-respiratory failure and shock | 4.8305 |
| Coronary atherosclerosis or angina | 0.05036 |
| Chronic obstructive pulmonary disease | 0.9216 |
| Decubitus ulcer or chronic skin ulcer | 0.3715 |
| Dementia, other specified brain disorders | 0.3746 |
| Diabetes mellitus or complications | 0.3503 |
| Dialysis status | 0.6628 |
| Hemiplegia, paraplegia, paralysis | 0.03503 |
| Major psychiatric disorders | –0.2813 |
| Metastatic cancer and acute leukemia | –0.4101 |
| Morbid obesity | 0.004568 |
| Osteoarthritis of hip or knee | –0.1722 |
| Bone/ cartilage disorders | –0.03929 |
| Other congenital deformity of hip/ knee | 0.05557 |
| Other major cancers | 0.3638 |
| Other injuries | –0.3110 |
| Pleural effusion/ pneumothorax | 0.2897 |
| Pneumonia | 1.9935 |
| Protein calorie malnutrition | 0.6445 |
| Renal failure | 0.06839 |
| Neoplasms | –0.1516 |
| Inflammatory connective tissue disease | 0.3672 |
| Stroke | 0.6596 |
| Trauma | –1.0302 |
| Vascular or circulatory disease | 0.6048 |
| Vertebral fractures, no spinal cord injury | 1.3644 |
| MGB, A | 0.3720 |
| MGB, B | 0.4111 |
| MGB, C | 0.3312 |
| MGB, D | 0.3278 |
| MGB, E | –0.1235 |
| MGB, F | –0.1533 |
| South U.S., A | –0.1585 |
| South U.S., B | –0.03940 |
| South U.S., C | –0.08401 |
| South U.S., D | –0.9982 |
| South U.S., E | –0.1733 |
| South U.S., F | –0.1126 |
| South U.S., G | 0.5828 |
| South U.S., H | 0.4366 |
| South U.S., I | –0.2882 |
| South U.S., J | –0.3628 |
| South U.S., K | 0.03241 |