

Machine Learning and Primary Total Knee Arthroplasty: Patient Forecasting for a Patient-Specific Payment Model

Sergio M. Navarro BS^{^*}
Eric Y. Wang BS^{*}
Heather S. Haeberle BS^{*}
Michael A Mont MD
Viktor E. Krebs MD[†]
Brendan M Patterson MD MBA[†]
Prem N. Ramkumar MD MBA[†]

[†] Cleveland Clinic, Department of Orthopedic Surgery, Cleveland, Ohio, USA

[^] Saïd Business School, University of Oxford, Oxford, UK

^{*} Baylor College of Medicine, Department of Orthopaedic Surgery, Houston, Texas, USA

Financial Disclosures: None

Figure should be published in color

Please send correspondence and requests for reprints to:

Sergio M. Navarro

Park End St, Oxford OX1 1HP, UK

Phone: +44 1865 270000

navarrosm@gmail.com

Development and Validation of a Machine-Learning Algorithm for Patient Forecasting after Primary Total Knee Arthroplasty in the Value-Based Era: A Patient-Specific Payment Model

Abstract

Background: Value-based and patient-specific care represent two critical areas of focus that have yet to be fully reconciled by today's bundled care model. Using a predictive naïve Bayesian model, the objectives of this study were: (1) to develop a machine-learning algorithm using preoperative big data to predict length of stay (LOS) and inpatient costs after primary TKA and (2) to propose a tiered patient-specific payment model that reflects patient complexity for reimbursement.

Methods: Using 141,446 patients undergoing primary TKA from an administrative database from 2009 to 2016, a Bayesian model was created and trained to forecast LOS and cost. Algorithm performance was determined using the area under the receiver operating characteristic curve (AUC) and the percent accuracy. A proposed risk-based patient-specific payment model was derived based on outputs.

Results: The machine-learning algorithm required age, race, gender, and comorbidity scores ("risk of illness" and "risk of morbidity") to demonstrate a high degree of validity with an AUC of 0.7822 and 0.7382 for LOS and cost. As patient complexity increased, cost add-ons increased in tiers of 3%, 10%, and 15% for moderate, major, and extreme mortality risks, respectively.

Conclusions: Our machine-learning algorithm derived from an administrative database demonstrated excellent validity in predicting LOS and costs prior to primary TKA and has broad value-based applications, including a risk-based patient-specific payment model.

Keywords: machine learning; total knee arthroplasty; predictive modeling

Machine Learning and Primary Total Knee Arthroplasty: Patient Forecasting for a Patient-Specific Payment Model

INTRODUCTION

The concept of value-based care has been critical for the development and implementation of alternative payment models (APMs) [1–3]. The Comprehensive Care for Joint Replacement (CJR) model of bundled payments and quality measures for compensation of lower extremity arthroplasty was established to improve value by incentivizing high quality care at lower costs [4–8]. Several institutions have demonstrated success in participation in the Bundled Payments for Care Improvement (BPCI) initiative for lower extremity joint arthroplasty [9–13]. The New York University Langone Medical Center demonstrated a 16% reduction in costs, with attributed savings in the post-acute care period by reducing utilization of inpatient rehabilitation facilities [12,14]. The Cleveland Clinic Euclid Hospital has reported similarly promising results, including a savings of over \$500,000 across 271 episodes of care, achieved by standardizing anesthetics with short-acting blocks and encouraging early post-operative mobilization [10]. With implementation of these changes, readmissions decreased from 5% to 1.6-2.7% and patients were almost twice as likely to be discharged to their home [10]. Hospitals that have demonstrated savings with bundled payments are more likely to be larger, have higher volumes, and be affiliated with post-acute care facilities [15]. While operations management strategies have been promising for controlling modifiable systemic risk factors related to inefficient care delivery, bundling care does not always take patient-specific risk factors into account.

Although systemic and procedural factors can have a direct impact on the success of bundled payment models, patient-level risk factors are integral in predicting costs and outcomes [16–19]. A single reimbursement fee for all total knee arthroplasty (TKA) procedures may be unsustainable in the long term, as it fails to take patient-level factors into account. This approach

may encourage “cherry-picking” younger, healthier patients, who are less likely to develop complications [20,21]. The current CJR reimbursement model stratifies patients by Diagnosis-Related Groups (DRGs) and geographic location, without accounting for other patient-related factors which may influence the costs of care [22–24]. In a survey of members of the American Association of Hip and Knee Surgeons (AAHKS), 94% of respondents were concerned that bundled payments would disincentivize surgeons from operating on higher risk patients [2]. Thus, a comprehensive model for identifying and quantifying patient-specific risk factors that influence outcomes and costs of delivering TKA is essential for determining appropriate reimbursements [25–28].

Machine learning has been increasingly applied to medical decision making, representing a natural extension of traditional statistical approaches [29–32]. Machine-learning processes exist along a spectrum from completely human-derived operations to fully automated programs, which process big data without human intervention [29,33]. A recent machine learning model produced the Risk of Inpatient Death, which uses 17 variables to estimate the risk of inpatient death for Intensive Care Unit (ICU) patients with 94% accuracy [30]. Deep machine learning has been applied to orthopaedic surgery, with Kim et al. demonstrating the ability to predict complications and mortality in posterior lumbar spine fusions [34]. These methods may be applied toward the advancement of value-based care in the setting of bundled payments in order to prospectively predict lengths of stay and payments of orthopaedic episodes-of-care based on patient-specific factors, as these models are ideally suited for stratifying patients into risk groups [35–37]. If machine learning models for the reimbursement of TKA demonstrate accuracy, there exists an opportunity to advance bundled payment models to account for individual variations, allowing for the development of a patient-specific payment model. This model maintains the

emphasis of bundled payment models on optimizing systemic factors while considering non-modifiable patient-level factors, appropriately compensating the care of complex patients, and mitigating the pressure to “cherry-pick” lower risk patients.

Therefore, the objectives of this study were: (1) to develop and validate a machine-learning model using preoperative big data to predict lengths of stay (LOS) and patient-specific inpatient payments after primary TKA for any given patient and (2) to propose a rudimentary patient-specific payment model that builds upon the bundled payment model through identification of patient complexity for reimbursement.

METHODS

In the first phase of the study, a machine-learning naïve Bayesian model was created from the New York State Department of Health’s Statewide Planning and Research Cooperative System (SPARCS). The model was subsequently validated in terms of systematic development, content validity, reliability, and responsiveness.

In the second phase of the study, we performed break-even analyses using the present non-tiered payment model to establish an evidence-based foundation for a tiered patient-specific payment model that adjusted the reimbursement and tested the model’s predictive accuracy for each tier.

Data Sources and Study Population

Data was gathered from the New York State Department of Health’s Statewide Planning and Research Cooperative System (SPARCS) database, a comprehensive reporting system, which collects patient-level data on all discharges from nonfederal acute-care hospitals in the state of New York. The dataset used included data from 141,446 patients between January 1, 2009 and December 31, 2016, as this date range encompassed the total patient-specific data available.

We used Clinical Classification Software (CCS) and All Patient Refined Diagnosis Related Groups (APR-DRGs) codes to define total knee arthroplasties. The cohort for the model included patients undergoing TKA that met the following criteria: (1) CCS Procedure Code 152 - “Arthroplasty knee” and (2) APR-DRG code - 302 “Knee Joint Replacement”. These criteria were used to ensure that the most comprehensive cohort of total knee arthroplasty candidates was incorporated. The aforementioned specifications were utilized to ensure that we captured patients

who underwent TKA, while excluding those patients who underwent revision TKA or partial knee arthroplasty (i.e., unicondylar arthroplasty).

Developing the Machine-learning Algorithm

All available data from the TKA surgical patients were extracted and analyzed for contribution to payments and LOS. The eight variables extracted were: age group, Charlson Comorbidity Score (CCS), ethnicity, gender, patient disposition, type of admission, All Patient Refined (APR) Risk of Mortality, and APR Severity of Illness. The CCS is based on comorbidities such as congestive heart failure, renal disease, and cancer documented from the 12 months before the hospitalization to 3 days after discharge [38]. The APR severity of illness is calculated using the DRG, specific to the calendar year of discharge [39]. Similarly, the APR risk of mortality is assigned using the APR-DRG and severity of illness [39]. Non-Medicare patients were excluded to standardize patients to a single insurance, Medicare, for ease of comparison. LOS was available within the database and defined as the number of days during the patient's operative admission. Payments were derived from the Institutional Cost Report (ICR) and New York SPARCS data. The ICR is a uniform report completed by New York State facilities to report income, expenses, assets, liabilities, and statistics to the Department of Health. Estimates of inpatient patients were calculated using hospital discharge data from the New York SPARCS and ICR data. ICR includes data on costs for each facility as well as ratios of Cost to Charges (RCCs). RCCs are certified, calculated, and reported by facilities and are subject to external audit. Specifically, the reported payment was the total charged for the patient's operative admission by the individual hospital less the attributed RCC amount. Thus, the payment was specific to both the orthopaedic procedure (i.e. total knee arthroplasty) and the hospital, since the

RCC, provided by ICR data, was multiplied by the charge data from the procedure, provided by the New York SPARCS data.

A Naïve Bayesian model classifier algorithm was created on a Lambda Quad graphics processing unit computer (Lambda Labs, San Francisco, California) with Intel (Santa Clara, California) Xenon Scalable processors and i7 5500U processors with 128GB of random access memory (RAM) using the Ubuntu operational system (Canonical, London, United Kingdom) [40,41]. A Python script was used to output the results script. For algorithm development, the LOS and Medicare payments of each patient undergoing primary TKA were computed. The build model forecasted LOS and payment after primary TKA using a 3:1 split in which 75% of the available patient data “built” the algorithm and the remaining 25% of patients were used for “testing.” During the initial training process, the parameters of the Naïve Bayesian model are initially set to the values of available patient data and using the eight patient parameters of the test set, the model supposes the LOS or payments [41]. For each incorrect prediction, the model self-calibrated with each future test patient based on review of possibly contributing patient factors that may affect payments or LOS [41]. This “learning” process was iterated five times for algorithm refinement, and the weighted contributions of each factor (“trained” data) were then applied to data for a “new” patient to predict post-operative LOS and payments. Five data sets were chosen for each available period of available data studied.

Evaluating the Algorithm

Model performance was determined using the area under the receiver operating characteristic curve (AUC) and percent accuracy, consistent with the prior work of Bradley et al. [42]. AUC is commonly used in classification analysis, including machine learning, in order to evaluate how well a model predicts classes. It is computed by evaluating the ROC curve and is

calculated to predict the probability that the classifier will rank a randomly chosen positive example greater than a randomly chosen negative example. As models provide a greater probability of fulfilling this prediction, the AUC approaches 1 [42]. Generally, an AUC of greater than 0.70 suggests a strong model [42]. Various types of cross validation for machine learning are used for testing and commonly accepted. The type of cross validation selected for this study was a "holdout method," which has been shown to perform better than traditional validation for machine learning. When the training and validation data sets vary considerably, the K-fold method is a viable validation strategy as it applies the hold-out method multiple times adjusting for bias and underfitting, but the lack of significant variability over the sets tested did not warrant this approach.

Inpatient payments were grouped as <\$10,000, \$10,000 to \$20,000, \$20,000 to \$25,000, \$25,000 to \$30,000, \$30,000 to \$40,000, and greater than \$40,000. Because 60% of the population fell between \$10,000 and 20,000, we used more granular groupings to capture extremes from the available data to capture increments of 5-10th percentiles. LOS was grouped as 1 to 3 days, 4 to 5 days, and greater than 5 days to similarly capture distribution extremes in terms of tertiles. Model validation was addressed in the context of construct validity, reliability, responsiveness, and systematic development [43]. Construct validity was able to be established through the use of the administrative database yielding pertinent outcome-specific measures of payment and LOS, as previously reported in several prior studies [43–46]. Responsiveness was established via the AUC curve, as described by de Vet et al., since this represents a combination of sensitivity and specificity for all possible cut-off change scores [47]. Reliability was established via the model accuracy across each of the five iterations as a different 3:1 split for each iteration, which showcases a reliable portrayal regardless of which population was used for

testing [47]. The systematic development of the model is inherent to the fact that the model identifies relevant items from preexisting studies undergoing primary TKA.

Risk-based Analysis

The payment model was applied to Medicare patients for each APR risk category (Minor, Moderate, Major, and Extreme). The predictive uncertainty error from the payment algorithm for moderate, major, and extreme in terms of payment were compared to those patients with a Minor APR risk. The error percentages were applied as an evidence-based proposal for a preliminary patient-specific payment model.

RESULTS

The Naïve Bayesian machine-learning algorithm was found to have an AUC of 0.7822 (95% CI 0.7796 - 0.7848) (**Figure 1**) and 0.7382 (95% CI 0.7355 - 0.7409) (**Figure 2**) after five iterations for LOS and payment, respectively. Groupings with the length of stay of greater than 5 days and payments of greater than \$30,000 resulted in the highest AUC respectively and minimized distance between any point on the ROC curve and the upper left intersection of the sensitivity and 1 – specificity values on the ROC graph. Progressive improvements after each iteration, or computational model “learning,” for LOS accuracy and AUC are depicted in **Table 1** and **Table 2**, respectively. **Table 3** and **Table 4** similarly depict payment AUC and accuracy, respectively. The average accuracy, or reliability, of the model was found to be 87.4% for LOS and 84.9% for payments after five data sets.

When stratified by APR Risk of Mortality, the predictive error, or risk, was increased by 3% for Moderate risk, 10% for Major risk, and 15% for Extreme risk when compared to Minor risk patients. After stratifying each of the actual payments to Medicare patients within the various payment strata, the mean Medicare inpatient payment were used as a base payment with the risk-based tiers then applied from the payment model to yield the proposed patient-specific payment model found in **Table 5**.

DISCUSSION

As the emphasis for delivering high-value care has increased, focus has shifted to improving the delivery of lower extremity arthroplasty, due to the high reimbursement by Medicare [48,49]. Although value has been traditionally characterized as a ratio of outcomes to cost, an underrecognized component of value is the individual patient [50]. Value is not standardized, as patients require different resources and have individual goals regarding outcomes. In order to apply this principle, value-based reimbursement models require a patient-centered focus in order to adjust for individual risk and provide personalized care.

Machine learning has expanded in its applications in recent years and is used daily to personalize user web browsing experiences, detect credit card fraud, and steer self-driving cars, but its practicality in medicine remains to be fully recognized [37,51]. As the implementation of machine learning algorithms in medicine continues to advance, processes will likely shift along the spectrum from human-derived operations toward machine-driven analyses [29]. Machine learning processes require extensive data sets, which are becoming increasingly available with the advent of big data [33,52,53]. These methods have potential applications in medicine in improving digital image identification, diagnostic accuracy, and prognostication [33,54]. Within orthopaedics, the ability to predict which patients are likely to require additional resources is vital to the success of APMs [1,2]. Models employing single reimbursement fees may encourage the inequitable practice of “cherry picking” total joint arthroplasty patients with fewer comorbidities in order to minimize financial risk [21]. With the predictive power of a machine-learning model, reimbursement of a bundled payment model is now capable of accounting for patient perioperative complexity through risk stratification. Using the data of over 140,000 patients undergoing primary TKA in New York State, we developed and validated a preliminary

framework for patient LOS and payment that demonstrated high responsiveness and reliability using AUC and accuracy, respectively. When the model was run specifically in terms of a comorbidity index, the model's predictive error for LOS and payment was used as a surrogate for "risk" and applied to propose a risk-based patient-specific payment model.

The primary objective of this study was to demonstrate that machine-learning methods provide accurate models for the prediction of a LOS and payments for individual patients, which may be applied preoperatively. With high accuracy and AUC values for both payment and LOS after five iterations, the Naïve Bayesian model demonstrated excellent responsiveness and reliability, respectively. This suggests that machine-learning algorithms are already capable of predicting LOS and inpatient costs during a primary TKA episode of care. The existing arthroplasty literature, both prospective and retrospective, has successfully identified risk factors for complications and payments in order to improve patient education and shared decision-making [27,55,56]. For example, the Risk Assessment and Predictor Tool, which stratifies patients based on age, sex, walking distance, use of a gait aid, use of community support, and presence of a caregiver at home, has been validated with overall predictive accuracy of 78% for discharge destination in total joint arthroplasty patients [57–59]. However, to date, no comprehensive model has been developed with the ability to preoperatively predict complications for total joint arthroplasty. This application of big data to the individual patient provides patients, surgeons, and payers with expectation management in order to participate in informed shared decision-making [55,60]. For example, diabetic patients with an HbA1c over 7.7% may benefit from preoperative expectation management counseling that they are at an increased risk of periprosthetic joint infection, and providers may provide more informed care [61]. Although individual risk factors have been previously identified and addressed, machine

learning provides the ability to incorporate these factors into a consolidated model for predicting the need for additional resources. With respect to payers, these models carry socio-political implications in the context of value-based care. Flat reimbursement fees financially incentivize providers and administrators to select for patients predisposed to the best outcomes, propagating access issues those with increased complexity.

The machine-learning model demonstrated that, while sensitivity and specificity as determined by the AUC was high for both models, inpatient payment AUC lagged slightly behind that of LOS by 0.044. This may be attributed to lack of standardized charges and inpatient practice variation; a larger cohort of patients across other institutions and insurance structures is required to assess this difference, as well as analysis before and after BPCI [10]. Models which accurately stratify patients by risk have many potential applications; this particular case provides a patient-specific payment model which accounts for patient risk and provides an evidence-based approach to building upon the existing bundled payments care model. Although the model demonstrates high responsiveness and reliability, *post hoc* analysis after five different data sets revealed a known predictive error range that increased incrementally for patients with higher APR Risk of Mortality scores. This represents the first data-driven approach to our knowledge that has proposed a risk-based, patient-specific payment model accounting for medical complexity in a primary TKA episode of care. A review of APMs by Siddiqi et al. demonstrated that costs of total joint arthroplasty may be reduced in bundled payments by reducing LOS, readmissions, and the utilization of inpatient rehabilitation facilities [48]. With a bundle reimbursement fee of approximately \$25,000 covering an entire inpatient and outpatient episode of primary TKA care, inpatient payments alone among extreme risk patients exceed \$25,000 65% and \$30,000 54% of the time (**Figure 3**). Thus, many hospitals may be unlikely to

assume the risk of performing primary TKA on this population even with maximal medical optimization in the flat-fee reimbursement model. As the population of older patients with additional comorbidities undergoing TKA grows, hospitals and surgeons will continue to bear the majority of the financial risk associated with flat-fee payment models [62]. The risk-stratified payments for the proposed patient-specific payment model encourage cost sharing, reduce patient selection, and reinforce patient access. If implemented, socio-economic disparities in the delivery of total joint arthroplasty may be lessened to encourage the health equity and quality of life at the population level.

This study is not without limitations. As a classifier algorithm, the Naïve Bayesian algorithm only forecasted which strata (i.e. 1 to 3 days, 4 to 5 days, and 6+ days) the tested patient was likely to belong, rather than providing a specific LOS or payment. Additionally, Naïve Bayesian algorithms are based on the conditional independence assumption that all attributes are independent, which fails to recognize cofounders between variables. Further along the spectrum of machine learning exist more complex models including deep learning neural networks, which may have the capability of providing more accurate models with specific outputs as opposed to stratifications [35]. Thus, future works may utilize these models to predict exact LOS or cost outputs rather than classify into strata. Additionally, machine learning models should ideally be tested for validation with an independent data set to avoid overestimating the accuracy of the model [33]. Furthermore, predictive machine learning models may need to be incrementally readjusted, as the effective half-life of the usefulness of historical data to predict future clinical performance may be as short as four months [63]. Despite these limitations, the Naïve Bayesian model represents the first critical step in establishing a machine-learning algorithm for improving value-based delivery of the elective TKA episode of care. In order to

validate a predictive machine-driven algorithm to derive a uniform patient-specific payment model, thorough investigation of a larger, more generalizable data set may be necessary. As such, further study is warranted to validate this algorithm in TKA. While the model was validated with a large administrative database in a single state, additional study and “learning” is required to strengthen the model using different databases to account for regional and population differences. The model was limited by the quality and content of the administrative database. Other factors, such as genetic predisposition to arthritis, prior history of dysplasia, socioeconomic status, and baseline functional demands, represents relevant attributes that warrant further exploration and may be addressed in future studies.

CONCLUSION

Our preliminary machine-learning algorithm derived from administrative big data demonstrated excellent construct validity, reliability, and responsiveness predicting LOS and payments prior to primary TKA. This has the potential to engender broad value-based applications, including a risk-based patient-specific payment model prior to elective TKA that offers tiered reimbursements commensurate with case complexity.

REFERENCES

- [1] Rozell JC, Courtney PM, Dattilo JR, Wu CH, Lee GC. Should All Patients Be Included in Alternative Payment Models for Primary Total Hip Arthroplasty and Total Knee Arthroplasty? *J Arthroplasty* 2016;31:45–9. doi:10.1016/j.arth.2016.03.020.
- [2] Kamath AF, Courtney PM, Bozic KJ, Mehta S, Parsley BS, Froimson MI. Bundled Payment in Total Joint Care: Survey of AAHKS Membership Attitudes and Experience with Alternative Payment Models. *J Arthroplasty* 2015;30:2045–56. doi:10.1016/j.arth.2015.05.036.
- [3] Wei DH, Hawker GA, Jevsevar DS, Bozic KJ. Improving value in musculoskeletal care delivery AOA critical issues. *J Bone Jt Surg - Am Vol* 2015;97:769–74. doi:10.2106/JBJS.N.00841.
- [4] McLawhorn AS, Buller LT. Bundled Payments in Total Joint Replacement: Keeping Our Care Affordable and High in Quality. *Curr Rev Musculoskelet Med* 2017;10:370–7. doi:10.1007/s12178-017-9423-6.
- [5] Maniya OZ, Mather RC, Attarian DE, Mistry B, Chopra A, Strickland M, et al. Modeling the Potential Economic Impact of the Medicare Comprehensive Care for Joint Replacement Episode-Based Payment Model. *J Arthroplasty* 2017;32:3268–3273.e4. doi:10.1016/j.arth.2017.05.054.
- [6] Halawi MJ, Greene K, Barsoum WK. Optimizing Outcomes of Total Joint Arthroplasty Under the Comprehensive Care for Joint Replacement Model. *Am J Orthop (Belle Mead NJ)* 2016;45:E112-3.
- [7] Chen KK, Harty JH, Bosco JA. It Is a Brave New World: Alternative Payment Models and Value Creation in Total Joint Arthroplasty: Creating Value for TJR, Quality and Cost-

313 Effectiveness Programs. *J Arthroplasty* 2017;32:1717–9. doi:10.1016/j.arth.2017.02.013.

314 [8] Navathe AS, Troxel AB, Liao JM, Nan N, Zhu J, Zhong W, et al. Cost of joint
315 replacement using bundled payment models. *JAMA Intern Med* 2017;177:214–22.
316 doi:10.1001/jamainternmed.2016.8263.

317 [9] Alfonso A, Hutzler L, Robb B, Beste C, Blom A, Bosco J. Similar Cost Savings of
318 Bundled Payment Initiatives Applied to Lower Extremity Total Joint Arthroplasty Can Be
319 Achieved Applying Both Models 2 and 3. *HSS J* 2017;13:267–70. doi:10.1007/s11420-
320 017-9571-7.

321 [10] Mouille B, Higuera C, Woicehovich L, Deadwiler M. How to Succeed in Bundled
322 Payments for Total Joint Replacement. *NEJM Catal* 2016.

323 [11] Edwards PK, Mears SC, Barnes CL. BPCI: Everyone Wins, Including the Patient. *J*
324 *Arthroplasty* 2017;32:1728–31. doi:10.1016/j.arth.2017.02.014.

325 [12] Jubelt LE, Goldfeld KS, Blecker SB, Chung WY, Bendo JA, Bosco JA, et al. Early
326 Lessons on Bundled Payment at an Academic Medical Center. *J Am Acad Orthop Surg*
327 2017;25:654–63. doi:10.5435/JAAOS-D-16-00626.

328 [13] Courtney PM, Ashley BS, Hume EL, Kamath AF. Are Bundled Payments a Viable
329 Reimbursement Model for Revision Total Joint Arthroplasty? *Clin Orthop Relat Res*
330 2016;474:2714–21. doi:10.1007/s11999-016-4953-6.

331 [14] Dummit LA, Kahvecioglu D, Marrufo G, Rajkumar R, Marshall J, Tan E, et al.
332 Association between hospital participation in a medicare bundled payment initiative and
333 payments and quality outcomes for lower extremity joint replacement episodes. *JAMA - J*
334 *Am Med Assoc* 2016;316:1267–78. doi:10.1001/jama.2016.12717.

335 [15] Navathe AS, Liao JM, Shah Y, Lyon Z, Chatterjee P, Polsky D, et al. Characteristics of

336 hospitals earning savings in the first year of mandatory bundled payment for hip and knee
 337 surgery. JAMA - J Am Med Assoc 2018;319:930–2. doi:10.1001/jama.2018.0678.

338 [16] Rondon AJ, Tan TL, Greenky MR, Goswami K, Shohat N, Phillips JL, et al. Who Goes to
 339 Inpatient Rehabilitation or Skilled Nursing Facilities Unexpectedly Following Total Knee
 340 Arthroplasty? J Arthroplasty 2018;33:1348–1351.e1. doi:10.1016/j.arth.2017.12.015.

341 [17] Courtney PM, Bohl DD, Lau EC, Ong KL, Jacobs JJ, Della Valle CJ. Risk Adjustment is
 342 Necessary in Medicare Bundled Payment Models for Total Hip and Knee Arthroplasty. J
 343 Arthroplasty 2018. doi:10.1016/j.arth.2018.02.095.

344 [18] Kurtz SM, Lau EC, Ong KL, Adler EM, Kolisek FR, Manley MT. Which Clinical and
 345 Patient Factors Influence the National Economic Burden of Hospital Readmissions After
 346 Total Joint Arthroplasty? Clin. Orthop. Relat. Res., vol. 475, 2017, p. 2926–37.
 347 doi:10.1007/s11999-017-5244-6.

348 [19] Saucedo JM, Marecek GS, Wanke TR, Lee J, Stulberg SD, Puri L. Understanding
 349 readmission after primary total hip and knee arthroplasty: Who’s at risk? J Arthroplasty
 350 2014;29:256–60. doi:10.1016/j.arth.2013.06.003.

351 [20] Clement RC, Derman PB, Kheir MM, Soo AE, Flynn DN, Levin LS, et al. Risk
 352 Adjustment for Medicare Total Knee Arthroplasty Bundled Payments. Orthopedics
 353 2016;39:e911–6. doi:10.3928/01477447-20160623-04.

354 [21] Humbyrd CJ. The Ethics of Bundled Payments in Total Joint Replacement: “Cherry
 355 Picking” and “Lemon Dropping.” J Clin Ethics 2018;29:62–8.

356 [22] Comprehensive Care for Joint Replacement Model. Centers Medicare Medicaid Serv 2018.
 357 <https://innovation.cms.gov/initiatives/cjr> (accessed May 28, 2018).

358 [23] Cairns MA, Moskal PT, Eskildsen SM, Ostrum RF, Clement RC. Are Medicare’s

359 “Comprehensive Care for Joint Replacement” Bundled Payments Stratifying Risk
360 Adequately? J Arthroplasty 2018. doi:10.1016/j.arth.2018.04.006.

361 [24] Ellimoottil C, Ryan AM, Hou H, Dupree J, Hallstrom B, Miller DC. Medicare’s new
362 bundled payment for joint replacement may penalize hospitals that treat medically
363 complex patients. Health Aff 2016;35:1651–7. doi:10.1377/hlthaff.2016.0263.

364 [25] Jain NB, Guller U, Pietrobon R, Bond TK, Higgins LD. Comorbidities increase
365 complication rates in patients having arthroplasty. Clin Orthop Relat Res 2005:232–8.
366 doi:10.1097/01.blo.0000156479.97488.a2.

367 [26] Zmistowski B, Karam JA, Durinka JB, Casper DS, Parvizi J. Periprosthetic Joint Infection
368 Increases the Risk of One-Year Mortality. J Bone Jt Surgery-American Vol
369 2013;95:2177–84. doi:10.2106/JBJS.L.00789.

370 [27] Ramkumar PN, Chu CT, Harris JD, Athiviraham A, Harrington MA, White DL, et al.
371 Causes and Rates of Unplanned Readmissions After Elective Primary Total Joint
372 Arthroplasty: A Systematic Review and Meta-Analysis. Am J Orthop (Belle Mead NJ)
373 2015;44:397–405.

374 [28] Sabeh KG, Rosas S, Buller LT, Freiberg AA, Emory CL, Roche MW. The Impact of
375 Medical Comorbidities on Primary Total Knee Arthroplasty Reimbursements. J Knee
376 Surg 2018. doi:10.1055/s-0038-1651529.

377 [29] Beam AL, Kohane IS. Big Data and Machine Learning in Health Care. JAMA
378 2018;319:1317. doi:10.1001/jama.2017.18391.

379 [30] Delahanty RJ, Kaufman D, Jones SS. Development and Evaluation of an Automated
380 Machine Learning Algorithm for In-Hospital Mortality Risk Adjustment Among Critical
381 Care Patients. Crit Care Med 2018;46:1. doi:10.1097/CCM.0000000000003011.

- [31] Char DS, Shah NH, Magnus D. Implementing Machine Learning in Health Care — Addressing Ethical Challenges. *N Engl J Med* 2018;378:981–3. doi:10.1056/NEJMp1714229.
- [32] Acharya UR, Oh SL, Hagiwara Y, Tan JH, Adeli H, Subha DP. Automated EEG-based screening of depression using deep convolutional neural network. *Comput Methods Programs Biomed* 2018;161:103–13. doi:10.1016/j.cmpb.2018.04.012.
- [33] Obermeyer Z, Emanuel EJ. Predicting the Future — Big Data, Machine Learning, and Clinical Medicine. *N Engl J Med* 2016;375:1216–9. doi:10.1056/NEJMp1606181.
- [34] Kim JS, Merrill RK, Arvind V, Kaji D, Pasik SD, Nwachukwu CC, et al. Examining the Ability of Artificial Neural Networks Machine Learning Models to Accurately Predict Complications Following Posterior Lumbar Spine Fusion. *Spine (Phila Pa 1976)* 2017;43:1. doi:10.1097/BRS.0000000000002442.
- [35] Ravi D, Wong C, Deligianni F, Berthelot M, Andreu-Perez J, Lo B, et al. Deep Learning for Health Informatics. *IEEE J Biomed Heal Informatics* 2017;21:4–21. doi:10.1109/JBHI.2016.2636665.
- [36] Ramkumar PN, Muschler GF, Spindler KP, Harris JD, McCulloch PC, Mont MA. Open mHealth Architecture: A Primer for Tomorrow’s Orthopedic Surgeon and Introduction to Its Use in Lower Extremity Arthroplasty. *J Arthroplasty* 2017;32:1058–62. doi:10.1016/j.arth.2016.11.019.
- [37] Chen JH, Asch SM. Machine Learning and Prediction in Medicine — Beyond the Peak of Inflated Expectations. *N Engl J Med* 2017;376:2507–9. doi:10.1056/NEJMp1702071.
- [38] Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *J Chronic*

405 Dis 1987;40:373–83. doi:10.1016/0021-9681(87)90171-8.

406 [39] New York State Department of Health. SPARCS Inpatient Output: Data Element
 407 Dictionary 2014:1–237.
 408 <https://www.health.ny.gov/statistics/sparcs/sysdoc/inpatientoutputdd.pdf> (accessed July 16,
 409 2018).

410 [40] Manning CS, Raghavan P, Schuetze H. Probabilistic information retrieval. An Introd. to
 411 Inf. Retr., Cambridge: Cambridge University Press; 2009, p. 219–35.

412 [41] Zhang H, Su J. Naive Bayesian Classifiers for Ranking. Univ New Brunswick n.d.

413 [42] Bradley AP. The use of the area under the ROC curve in the evaluation of machine
 414 learning algorithms. Pattern Recognit 1997;30:1145–59. doi:10.1016/S0031-
 415 3203(96)00142-2.

416 [43] Ramkumar PN, Harris JD, Noble PC. Patient-reported outcome measures after total knee
 417 arthroplasty: a systematic review. Bone Jt Res 2015;4:120–7. doi:10.1302/2046-
 418 3758.47.2000380.

419 [44] Naal FD, Impellizzeri FM, Sieverding M, Loibl M, von Knoch F, Mannion AF, et al. The
 420 12-item Oxford Knee Score: cross-cultural adaptation into German and assessment of its
 421 psychometric properties in patients with osteoarthritis of the knee. Osteoarthr Cartil
 422 2009;17:49–52. doi:10.1016/j.joca.2008.05.017.

423 [45] Ko Y, Lo NN, Yeo SJ, Yang KY, Yeo W, Chong HC, et al. Rasch analysis of the Oxford
 424 Knee Score. Osteoarthr Cartil 2009;17:1163–9. doi:10.1016/j.joca.2009.04.004.

425 [46] Bershadsky B, Kane RL, Wuerz T, Jones M, Brighton B, Stitzlein R, et al. Preliminary
 426 validation of the Review of Musculoskeletal System (ROMS) questionnaire. J Bone Jt
 427 Surg - Am Vol 2015;97:582–9. doi:10.2106/JBJS.M.01078.

428 [47] de Vet HCW, Bouter LM, Bezemer PD, Beurskens AJHM. Reproducibility and
429 responsiveness of evaluative outcome measures. Theoretical considerations illustrated by
430 an empirical example. *Int J Technol Assess Health Care* 2001;17:479–87.

431 [48] Siddiqi A, White PB, Mistry JB, Gwam CU, Nace J, Mont MA, et al. Effect of Bundled
432 Payments and Health Care Reform as Alternative Payment Models in Total Joint
433 Arthroplasty: A Clinical Review. *J Arthroplasty* 2017;32:2590–7.
434 doi:10.1016/j.arth.2017.03.027.

435 [49] Piccinin MA, Sayeed Z, Kozlowski R, Bobba V, Knesek D, Frush T. Bundle Payment for
436 Musculoskeletal Care: Current Evidence (Part 2). *Orthop Clin North Am* 2018;49:147–56.
437 doi:10.1016/j.ocl.2017.11.003.

438 [50] Bozic KJ, Wright JG. Value-based healthcare and orthopaedic surgery: Editorial comment.
439 *Clin Orthop Relat Res* 2012;470:1004–5. doi:10.1007/s11999-012-2267-x.

440 [51] Darcy AM, Louie AK, Roberts LW. Machine learning and the profession of medicine.
441 *JAMA - J Am Med Assoc* 2016;315:551–2. doi:10.1001/jama.2015.18421.

442 [52] Halevy A, Norvig P, Pereira F. The Unreasonable Effectiveness of Data. *IEEE Intell Syst*
443 2009;24:8–12. doi:10.1109/MIS.2009.36.

444 [53] Krumholz HM. Big data and new knowledge in medicine: The thinking , training , and
445 tools needed for a learning health system. *Health Aff* 2014;33:1163–70.
446 doi:10.1377/hlthaff.2014.0053.

447 [54] Lee TH, Cosgrove T. Engaging Doctors in the Healthcare Revolution. *Harv Bus Rev*
448 2014;92:105–11.

449 [55] Arden N, Altman D, Beard D, Carr A, Clarke N, Collins G, et al. Lower limb arthroplasty:
450 can we produce a tool to predict outcome and failure, and is it cost-effective? An

- epidemiological study. *Program Grants Appl Res* 2017;5:1–246. doi:10.3310/pgfar05120.
- [56] Courtney PM, Rozell JC, Melnic CM, Lee GC. Who Should Not Undergo Short Stay Hip and Knee Arthroplasty? Risk Factors Associated With Major Medical Complications Following Primary Total Joint Arthroplasty. *J Arthroplasty* 2015;30:1–4. doi:10.1016/j.arth.2015.01.056.
- [57] Hansen VJ, Gromov K, Lebrun LM, Rubash HE, Malchau H, Freiberg AA. Does the Risk Assessment and Prediction Tool Predict Discharge Disposition After Joint Replacement? *Clin Orthop Relat Res* 2014;473:597–601. doi:10.1007/s11999-014-3851-z.
- [58] Slover J, Mullaly K, Karia R, Bendo J, Ursomanno P, Galloway A, et al. The use of the Risk Assessment and Prediction Tool in surgical patients in a bundled payment program. *Int J Surg* 2017;38:119–22. doi:10.1016/j.ijssu.2016.12.038.
- [59] Dauty M, Schmitt X, Menu P, Rousseau B, Dubois C. Using the Risk Assessment and Predictor Tool (RAPT) for patients after total knee replacement surgery. *Ann Phys Rehabil Med* 2012;55:4–15. doi:10.1016/j.rehab.2011.10.006.
- [60] Collins FS, Varmus H. A New Initiative on Precision Medicine. *N Engl J Med* 2015;372:793–5. doi:10.1056/NEJMp1500523.
- [61] Tarabichi M, Shohat N, Kheir MM, Adelani M, Brigati D, Kearns SM, et al. Determining the Threshold for HbA1c as a Predictor for Adverse Outcomes After Total Joint Arthroplasty: A Multicenter, Retrospective Study. *J Arthroplasty* 2017;32:S263–S267.e1. doi:10.1016/j.arth.2017.04.065.
- [62] Rubin LE, Blood TD, Defillo-Draiby JC. Total Hip and Knee Arthroplasty in Patients Older Than Age 80 Years. *J Am Acad Orthop Surg* 2016;24:683–90. doi:10.5435/JAAOS-D-15-00278.

474 [63] Chen JH, Alagappan M, Goldstein MK, Asch SM, Altman RB. Decaying relevance of
475 clinical data towards future decisions in data-driven inpatient clinical order sets. *Int J Med*
476 *Inform* 2017;102:71–9. doi:10.1016/j.ijmedinf.2017.03.006.
477
478
479

TABLES AND FIGURES:

Figure 1. AUC curve analyzing predictive accuracy of Bayesian model for LOS.

Figure 2. AUC curve analyzing predictive accuracy of Bayesian model for cost.

Figure 3. Distribution of inpatient payments stratified by APR Risk of Mortality relative to six cost strata

Table 1. Depiction of naïve Bayesian model for each of the five iterations of LOS and accuracy

Table 2. Depiction of naïve Bayesian model for each of the five iterations of LOS and AUC

Table 3. Depiction of naïve Bayesian model for each of the five iterations of payment and accuracy

Table 4. Depiction of naïve Bayesian model for each of the five iterations of payment and AUC

Table 5. Proposed patient-specific payment model using model's predictive error

Figure 1
[Click here to download high resolution image](#)

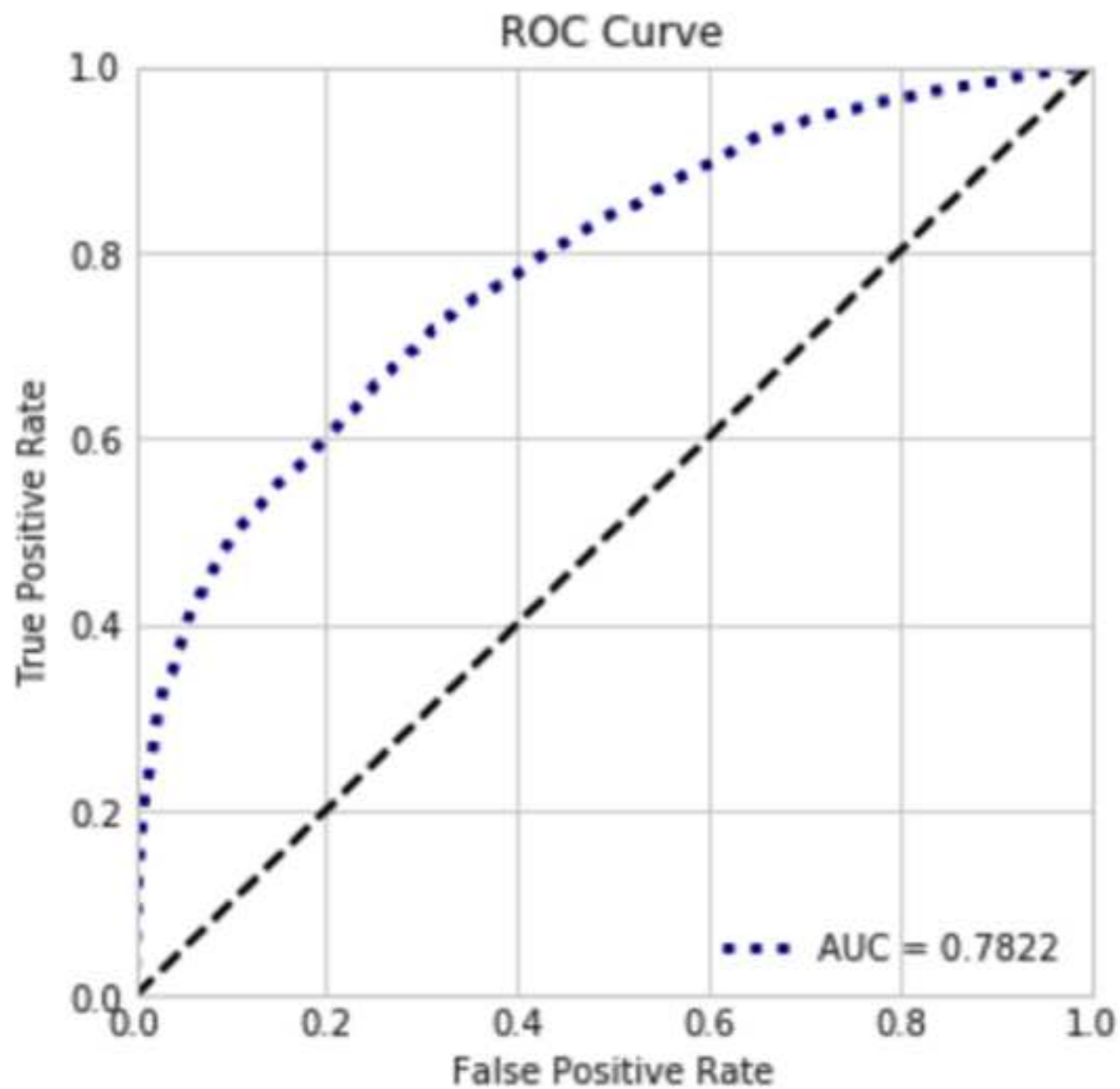


Figure 2
[Click here to download high resolution image](#)

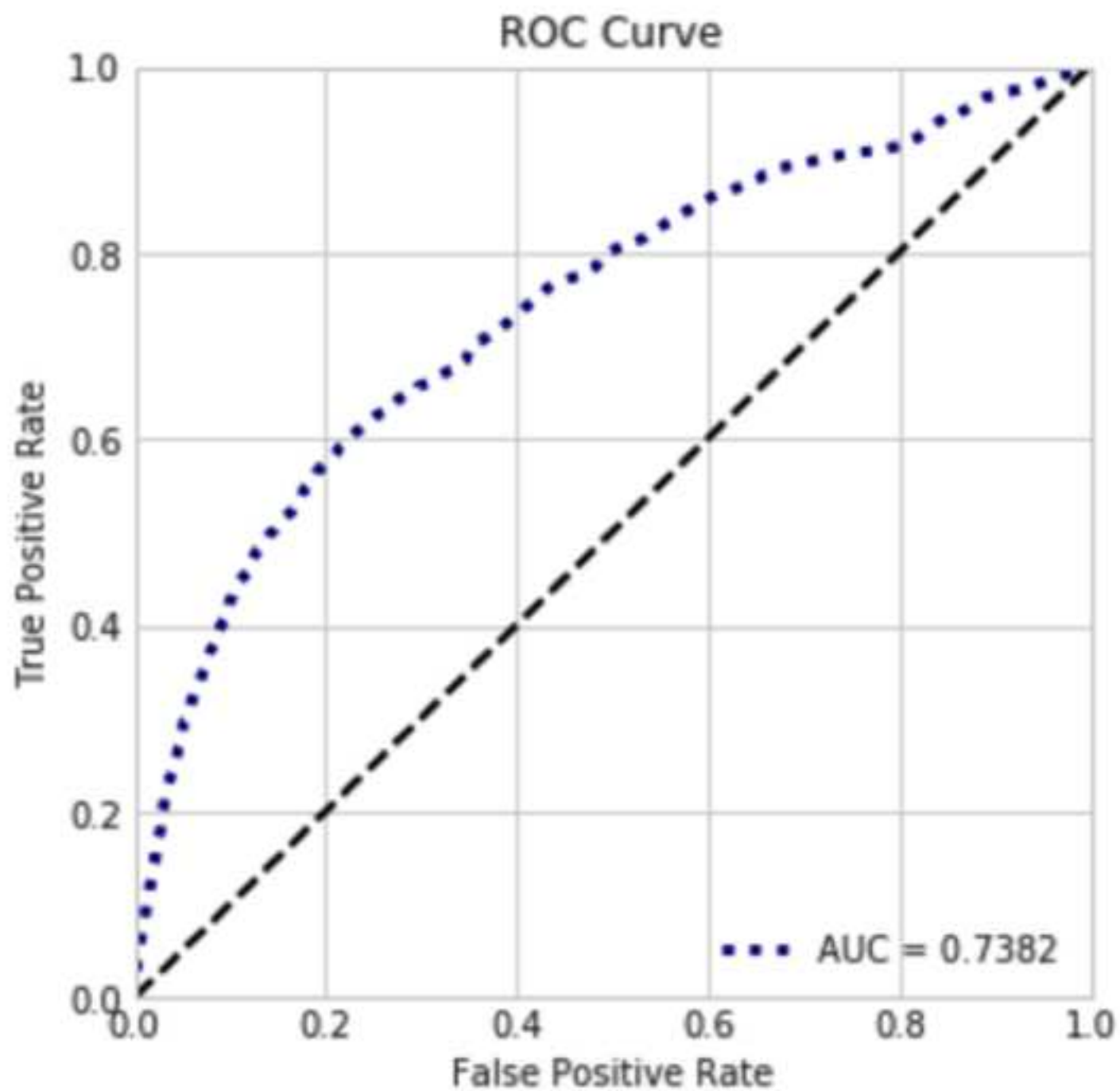
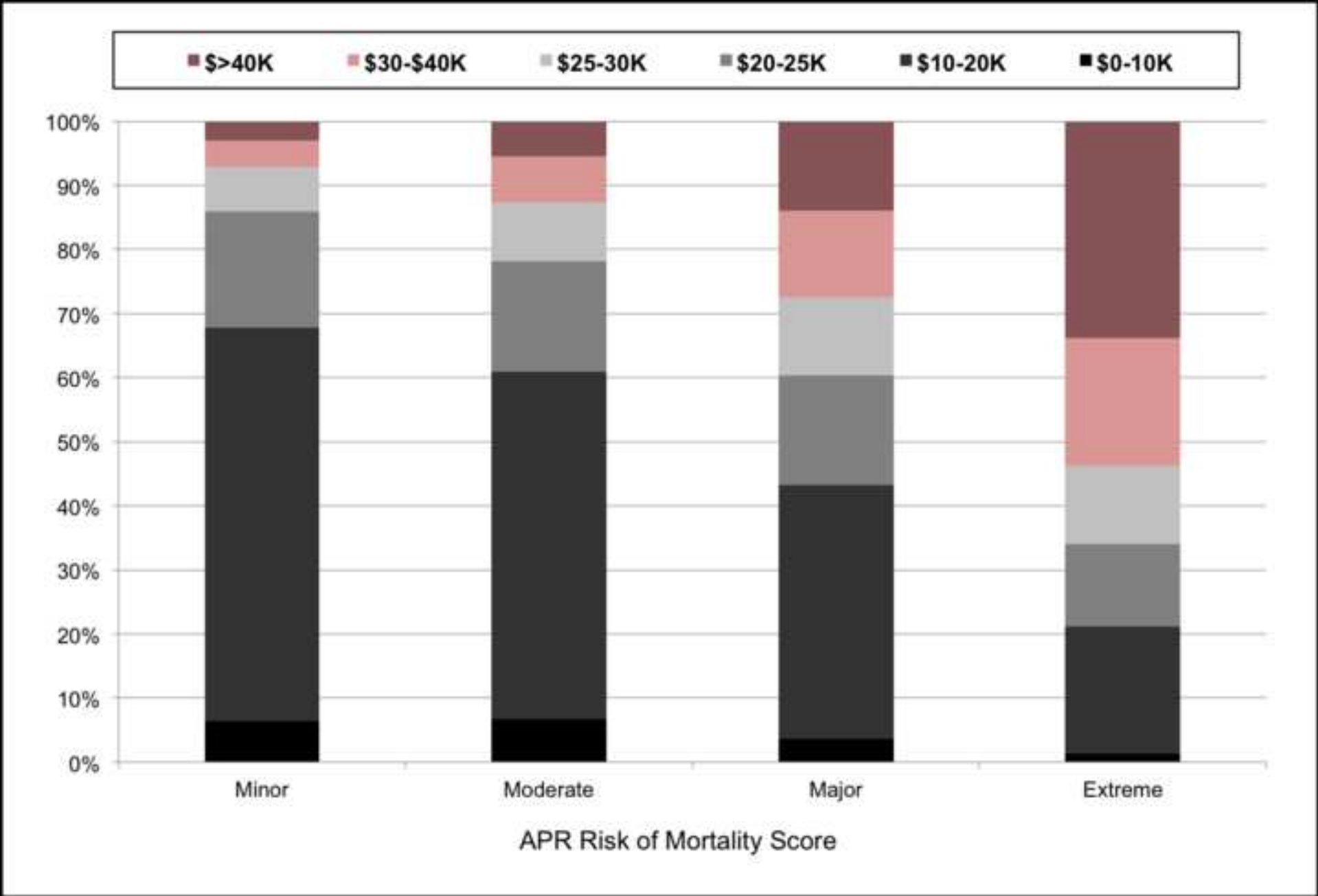


Figure 3
[Click here to download high resolution image](#)



FIGURES:

Figure 1. AUC curve analyzing predictive accuracy of Bayesian model for LOS.

Figure 2. AUC curve analyzing predictive accuracy of Bayesian model for cost.

Figure 3. Distribution of inpatient costs stratified by APR Risk of Mortality relative to six cost strata

Table 1

Table 1. Depiction of naïve Bayesian model for each of the five iterations of LOS and accuracy

Data Sets	LOS Accuracy
	(1-3, 4-5, >5 days)
Set 1	0.849955869
Set 2	0.863770686
Set 3	0.878372681
Set 4	0.885328537
Set 5	0.892003117
Average	0.873886178

Table 3

Table 3. Depiction of naïve Bayesian model for each of the five iterations of cost and accuracy

Data Sets	Cost Accuracy (<\$10K, \$10K-\$20K, \$20K -\$25K , \$25K-\$30K, \$30K-\$40K, >\$40K)
Set 1	0.819947043
Set 2	0.837201805
Set 3	0.853235666
Set 4	0.861593799
Set 5	0.870988165
Average	0.848668664

Table 2. Depiction of naïve Bayesian model for each of the five iterations of LOS and AUC

	LOS AUC	
Data Sets	Average (>4 days)	Average (>5 days)
Set 1	0.679808	0.715495
Set 2	0.715588	0.753152
Set 3	0.753250	0.792792
Set 4	0.768315	0.808648
Set 5	0.799048	0.840994
Average	0.743200	0.782200

Table 4. Depiction of naïve Bayesian model for each of the five iterations of cost and AUC

	Cost AUC	
Data Sets	Average (>\$25K)	Average (>\$30K)
Set 1	0.64514	0.67523
Set 2	0.67909	0.71077
Set 3	0.71483	0.74818
Set 4	0.72913	0.76314
Set 5	0.75830	0.79367
Average	0.70530	0.73820

Table 5. Proposed patient-specific payment model using model's predictive error based on APR
Risk of Mortality

Iterations	Risk	Mean Cost	Risk-adjust
Minor	100%	18,100	18,100.0
Moderate	103%	19,200	19,700.0
Major	110%	22,700	25,000.0
Extreme	115%	30,100	34,600.0